



Dissertação

Mestrado em Gestão de Sistemas de Informação Médica

Conversão de objetos 3D em sons tridimensionais

Nuno Miguel Lourenço Almeida

Leiria, setembro de 2018



Dissertação

Mestrado em Gestão de Sistemas de Informação Médica

Conversão de objetos 3D em sons tridimensionais

Nuno Miguel Lourenço Almeida

Dissertação de Mestrado realizada sob a orientação do Doutor João Pereira, Professor do Departamento de Engenharia Informática da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria.

Leiria, *setembro* de 2018

À Minha Família

Esta página foi intencionalmente deixada em branco

Agradecimentos

Em primeiro lugar quero agradecer à minha família por todo o apoio nesta caminhada e por me proporcionarem esta oportunidade. Um especial agradecimento à minha mãe por tudo o que faz por mim.

Quero agradecer também aos meus amigos que sempre tiveram uma palavra de motivação durante esta caminhada.

Agradeço ao meu orientador Prof. João Pereira que sempre se mostrou disponível para ajudar na realização deste trabalho.

Por último agradeço a todas as pessoas que de uma forma ou de outra me apoiaram e auxiliaram nesta etapa académica.

Esta página foi intencionalmente deixada em branco

Resumo

Esta dissertação tem como objetivo apresentar um sistema capaz de converter a forma de um objeto ou de um cenário tridimensional em som com o objetivo de auxiliar pessoas com sérias dificuldades visuais. Com o intuito de contextualizar este trabalho começa-se por fazer um enquadramento relativo a pessoas com dificuldades visuais. São apresentados também os potenciais benefícios resultantes deste trabalho. De forma a dar suporte a este trabalho, na revisão de literatura é feito o enquadramento relativo a efeitos de áudio 3D, apresentam-se soluções já existentes no mercado e que de alguma forma se assemelham ou fornecem uma possível alternativa a uma parte da solução proposta nesta dissertação. São ainda descritas as tecnologias utilizadas nesta dissertação. O protótipo desenvolvido neste trabalho distancia-se de outras soluções de conversão de imagem em som uma vez que o som é gerado com base num objeto 3D em vez de imagens 2D, fornecendo assim a informação relativa à forma e profundidade de um objeto de forma mais precisa. A digitalização de um objeto permite fazer a sua divisão em camadas representadas por diversas fontes sonoras virtuais. Estas camadas são percorridas da mais próxima à mais distante. Para cada camada é efetuado o cálculo da distância entre o utilizador e as fontes sonoras de forma a gerar um som específico para os 6 pares de auscultadores, sendo a frequência dos sons diferentes para cada camada. Pelos resultados observados, é possível concluir que a digitalização de um objeto e a sua conversão em som permite fazer a identificação de diferentes objetos.

Palavras-chave: objetos 3D, sons tridimensionais, scanner 3D, headphones 3D, dificuldades visuais

Esta página foi intencionalmente deixada em branco

Abstract

This dissertation aims to present a system capable of converting the shape of an object or a three-dimensional scenery into sound in order to assist people with serious visual impairment. In order to contextualize this work, it starts by making an overview of the difficulties visually impaired people have to deal with. Also presented are the potential benefits resulting from this work. In order to support this work, in the literature review, an overview of 3D audio effects is made, some solutions already existing in the market and that somehow resemble or provide a possible alternative to a part of the proposed solution in this dissertation are presented. Also described are the technologies used in this dissertation. The prototype developed in this work distances itself from other solutions of image to sound conversion since the sound is generated based on a 3D object instead of 2D images, thus providing information regarding the shape and depth of an object with an increased level of precision. The digitization of an object allows its division into layers represented by several virtual sound sources. These layers are traversed from the nearest to the farthest. For each layer, the distance between the user and the sound sources is calculated in order to generate a specific sound for the 3 pairs of speakers, with the frequency of sounds being different for each layer. From the observed results, it is possible to conclude that the scanning of an object and its conversion into sound allows the identification of different objects.

Key-Words: 3D objects, 3D sounds, 3D scanner, 3D headphones, visually impaired

Esta página foi intencionalmente deixada em branco

Índice de Figuras

Figura 1 - Decomposição de um objeto 3D, por camada, no sistema Cartesiano	15
Figura 2 - Representação de um dos lados dos headphones	17
Figura 3 - <i>Headphones</i> impressos.....	17
Figura 4 - Representação de uma camada, dos seus pontos e distâncias entre os auscultadores..	19
Figura 5 - Diagrama do funcionamento do protótipo	19
Figura 6 – Conversão de uma esfera	20
Figura 7 - Porta fechada	21
Figura 8 - Porta semiaberta	22
Figura 9 - Pessoa com o baço levantado	23

Esta página foi intencionalmente deixada em branco

Índice de Tabelas

Tabela 1 - Resultados testes utilizadores reais	24
---	----

Esta página foi intencionalmente deixada em branco

Lista de Siglas

2D	Bidimensional
3D	Tridimensional
DB	Direito Baixo
DC	Direito Cima
DT	Direito Traseiro
EB	Esquerdo Baixo
EC	Esquerdo Cima
ET	Esquerdo Traseiro
HRTF	<i>Head-Related Transfer Function</i>
LabVIEW	<i>Laboratory Virtual Instrument Engineering Workbench</i>
LED	<i>Light-emited diode</i>
RGB	<i>Red, Green and Blue</i>
SDK	<i>Software Developer's Kit</i>
STL	<i>StereoLithography</i>

Esta página foi intencionalmente deixada em branco

Índice

DEDICATÓRIA.....	I
CONVERSÃO DE OBJETOS 3D EM SONS TRIDIMENSIONAIS	I
AGRADECIMENTOS	IV
RESUMO.....	VI
ABSTRACT	VIII
ÍNDICE DE FIGURAS	X
ÍNDICE DE TABELAS	XII
LISTA DE SIGLAS	XIV
ÍNDICE	XVI
1 INTRODUÇÃO.....	1
1.1 CONTEXTO E MOTIVAÇÃO.....	1
1.2 POTENCIALIDADES E CONTRIBUTOS	2
1.3 OBJETIVOS	2
1.4 ORGANIZAÇÃO DO DOCUMENTO.....	2
2 REVISÃO DA LITERATURA.....	3
2.1 ÁUDIO 3D.....	3
2.1.1 <i>Áudio binaural</i>	3
2.1.2 <i>HRTF</i>	3
2.2 SOLUÇÕES EXISTENTES	4
2.2.1 <i>vOICe</i>	4
2.2.2 <i>Blue Edge Bulgaria</i>	5
2.2.3 <i>Assisting the visually impaired</i>	5
2.2.4 <i>Seeing the Movement through Sound</i>	6
2.2.5 <i>Seeing the Sound</i>	7
2.2.6 <i>Scanner 3D de imagens parciais</i>	8
2.3 <i>MICROSOFT KINECT</i>	8
2.4 <i>LABVIEW</i>	9
2.5 <i>FORMATO STL</i>	10
3 METODOLOGIA	11
3.1 PROBLEMA	11
3.2 PROCESSO DE INVESTIGAÇÃO	11
3.3 SELEÇÃO DAS FERRAMENTAS A USAR	12
3.4 DESENVOLVIMENTO DO PROTÓTIPO	13
4 SOLUÇÃO PROPOSTA	15
4.1 DESCRIÇÃO TEÓRICA DO PROCESSO	15
4.2 IDENTIFICAÇÃO DO <i>HARDWARE</i> UTILIZADO	16
4.3 MODO DE FUNCIONAMENTO DO PROTÓTIPO.....	17
5 DISCUSSÃO DE RESULTADOS.....	20
5.1 RESULTADOS DA CONVERSÃO DE OBJETOS EM SOM	20
5.2 UTILIZADORES REAIS.....	23

6	CONCLUSÃO	25
6.1	SÍNTESE DO TRABALHO REALIZADO.....	25
6.2	RESULTADOS.....	25
6.3	LIMITAÇÕES	26
6.4	TRABALHO FUTURO	26
	BIBLIOGRAFIA	27
	APÊNDICES.....	28
	ANEXOS	29
	ANEXO 1 – PATENTE NACIONAL N.º WO 2017/221177 A1	30
	ANEXO 2 – ARTIGO PUBLICADO NA CONFERÊNCIA “INDIA’S DIGITAL TRANSFORMATION”	53
	ANEXO 3 - ARTIGO ACEITE NA CONFERÊNCIA “IEEE SPONSORED 4TH INTERNATIONAL CONFERENCE FOR CONVERGENCE IN TECHNOLOGY (I2CT) 2018”	63

Esta página foi intencionalmente deixada em branco

1 Introdução

Atualmente estima-se que existam cerca de 253 milhões de pessoas com algum tipo de dificuldade visual. Estas dificuldades podem-se manifestar em qualquer pessoa independentemente da origem educação ou status social da pessoa. Nos casos mais graves, estas dificuldades podem representar a necessidade de contato físico para detetar ou reconhecer objetos.

É neste sentido que surge a motivação para a elaboração da dissertação de Mestrado: desenvolver um sistema capaz de captar um cenário tridimensional e fazer a sua representação sonora.

De seguida serão apresentados os motivos para a criação do sistema.

1.1 Contexto e motivação

Segundo a Organização Mundial de Saúde, estima-se que cerca de 36 milhões de pessoas sofram de cegueira e que 217 milhões têm comprometimento moderado ou grave da visão. Pessoas com 50 ou mais anos compõem 81% das pessoas que têm comprometimento moderado ou severo de visão ou cegueira. O envelhecimento da população faz com que, cada vez mais, haja aumento do número de pessoas em risco de perder alguma visão devido a doenças oculares crónicas. No caso de jovens com idades inferiores a 15 anos, estima-se que cerca de 1,4 milhões sofram de cegueira irreversível, necessitando de algum tipo de serviço que as ajude a completar tarefas do dia-a-dia. [1]

Os sistemas atualmente existentes que podem servir de ajuda para estas pessoas incidem maioritariamente na deteção ou localização de objetos ou baseiam-se apenas em imagens 2D para fazer o reconhecimento do aspeto de um objeto. Desta forma existe a motivação de criar um sistema capaz de fornecer a pessoas com dificuldades visuais uma forma de fazer o reconhecimento de objetos com um elevado nível de precisão.

1.2 Potencialidades e Contributos

No sentido de contribuir para o melhoramento do nível de vida de pessoas com dificuldades visuais é necessário criar mecanismos que ajudem não só a detetar a existência de objetos, mas também que seja possível fornecer ao utilizador uma forma de identificar esses mesmos objetos.

Esta dissertação acarreta potencial pelo facto de apresentar um sistema capaz de fornecer ao utilizador uma representação sonora da forma aproximada de um objeto ou cenário, permitindo assim que este não só detete, mas também identifique objetos/obstáculos.

Todas as potencialidades mencionadas acima promovem os objetivos deste trabalho, que são apresentados na secção seguinte.

1.3 Objetivos

Os sistemas que existem atualmente baseiam-se na conversão de uma única imagem 2D em som ou limitam-se a fazer a deteção e localização de obstáculos. De forma a que seja possível fornecer uma informação mais detalhada dos objetos e dessa forma ajudar a melhorar a qualidade de vida das pessoas com dificuldades visuais, o principal objetivo desta dissertação é a criação de um sistema que seja capaz de captar um objeto ou cenário tridimensional e fazer a sua representação utilizando som.

No capítulo seguinte será descrita a organização do presente documento e dos tópicos de cada capítulo.

1.4 Organização do documento

A dissertação inicia-se com a introdução (capítulo 1) ao contexto atual das pessoas com dificuldades auditivas no mundo. No capítulo 2 é feita a revisão da literatura, secção onde serão expostos alguns conceitos de áudio tridimensional, trabalhos já desenvolvidos na área e o enquadramento das ferramentas utilizadas. De seguida no capítulo 3º é apresentada a metodologia seguida para a realização desta dissertação. No capítulo 4 apresenta-se a solução proposta resultante do trabalho desenvolvido ao longo desta dissertação, seguindo-se a discussão de resultados sobre o trabalho realizado (capítulo 5), e, por fim, no capítulo 6 são descritas as conclusões da dissertação bem como propostas para trabalho futuro.

2 Revisão da literatura

Neste capítulo são apresentadas técnicas de áudio 3D, seguindo-se a apresentação de soluções existentes para a conversão de imagem em som e por fim são apresentadas as tecnologias usadas nesta dissertação

2.1 Áudio 3D

Efeitos de áudio 3D são um conjunto de efeitos sonoros que manipulam o som proveniente de diversos tipos de colunas. Isto envolve a colocação virtual de fontes sonoras num espaço tridimensional, incluindo por cima, por baixo e por trás do utilizador. Estes efeitos transformam ondas sonoras de forma a imitar ondas sonoras naturais provenientes de uma fonte num espaço tridimensional.

Estes efeitos recorrem maioritariamente a HRTF (*head-related transfer functions*) para simular as alterações que o som sofre desde a sua emissão até chegar aos ouvidos do utilizador. Algumas soluções recorrem também a gravações de áudio binaural.

2.1.1 Áudio binaural

Áudio binaural é áudio captado de forma idêntica à forma como os ouvidos captam o som. Este áudio é capturado com recurso a 2 microfones colocados a uma distância de 18 cm entre si, de forma a simular a localização dos ouvidos.

O áudio binaural não deve ser confundido com o som stereo tradicional. Os sons stereo apenas fazem a distinção entre esquerda e direita, enquanto o som binaural permite ao utilizador localizar as fontes de som à sua frente, por trás de si, em cima e em baixo. Outra diferença entre o som stereo e binaural é o fato de o som stereo não considerar a localização dos ouvidos.[2]

2.1.2 HRTF

Uma HRTF é uma resposta que caracteriza a forma como uma orelha recebe um estímulo sonoro proveniente de um ponto no espaço. Este estímulo é influenciado não só pela estrutura da cabeça humana, mas também pelo resto do corpo. Qualquer obstáculo que se encontre entre a emissão

de um sinal sonoro e o seu destinatário irá alterar a frequência e a fase desse mesmo sinal. O cérebro humano está preparado para processar e entender estes sinais de forma a determinar a localização da origem do sinal sonoro. De uma forma resumida pode-se dizer que as HRTF descrevem o impacto que a anatomia de cada pessoa tem no som proveniente de qualquer localização.

A capacidade de uma pessoa localizar a origem de um som provém da análise que o cérebro humano faz dos atributos do som. Um dos atributos mais importantes é a diferença do som na orelha direita e na orelha esquerda. A diferença de tempo que um som demora para chegar a cada uma das orelhas reflete-se na diminuição da amplitude da onda que chega à orelha que se encontra mais distante da origem do som.

Um grande obstáculo das HRTF é o fato de estas serem computacionalmente exigentes. Para além disso, uma vez que cada pessoa tem um formato de orelha diferente, as HRTF necessitam de ser *customizadas* para cada utilizador. Para tal é necessária tecnologia capaz de gerar HRTF personalizadas, o que implica a criação de bases de dados de formatos de orelhas.

2.2 Soluções existentes

Nesta secção serão apresentadas algumas soluções existentes para a conversão de imagem em som e descrito o seu modo de funcionamento. Também serão apresentados alguns *scanners* 3D.

2.2.1 vOICe

Em 1992, Peter Meijer apresentou o vOICe[3], um sistema capaz de traduzir com som as imagens de um vídeo captado por uma câmara. A capacidade para uma pessoa aprender a interpretar o som da imagem depende das capacidades dessa mesma pessoa.

Este sistema efetua um varrimento por segundo de uma imagem com uma resolução de 64×64 da esquerda para a direita, associando a elevação da imagem com tom e brilho com volume. Cada pixel numa coluna produz uma onda cuja frequência indica a sua posição, encontrando-se as frequências mais altas no topo. As ondas sonoras são geradas com base numa escala de cinzas de 16 tons. Uma vez extraída toda a informação da imagem, o sistema selecciona o *frame* seguinte.

De forma a aumentar a orientação espacial do utilizador, são usados uns *headphones stereo* que alternam o volume da esquerda para a direita em consonância com o pixel que está a ser processado. Este efeito permite que o utilizador identifique a localização de um objeto. Apesar da simplicidade deste processo, os sons gerados são geralmente complexos devido à variedade de informação presente na imagem.

Originalmente foi desenvolvido um computador específico ligado a uma câmara de vídeo de baixa resolução, mas versões mais recentes utilizam computadores normais e webcams ou dispositivos com maior portabilidade como um *Raspberry Pi* ou um dispositivo *Android*.

A versão mais recente deste sistema é capaz de gerar “paisagens sonoras” com uma resolução de 176×64 e permite ouvir portas, edifícios, árvores, entre outros. Também pode ser usado como sistema de aviso para ajudar pessoas com perda parcial ou total de visão a detetar alterações na periferia visual.

Este sistema utiliza de forma parcial uma invenção que permite a transformação de qualquer informação que possa ser representada no formato $M \times N$ numa representação sonora. Cada pedaço de informação possui propriedades específicas, tais como a intensidade do pixel em questão.

Este sistema serviu de base para diversos outros sistemas de conversão de imagem em áudio.

2.2.2 Blue Edge Bulgaria

Em outubro de 2003, a Blue Edge Bulgaria anunciou o desenvolvimento de um *software* capaz de transformar um conjunto de telemóveis em dispositivos de auxílio visual através da transformação das imagens captadas pela câmara do telemóvel em sons que permitissem ao cérebro reconstruir essas imagens com base no som.

Este sistema é uma implementação do sistema vOICE num telemóvel. A grande diferença desta implementação para o sistema original prende-se com o fato de utilizar imagens estáticas em vez de imagens em movimento.

2.2.3 Assisting the visually impaired

Em 2012 foi apresentado um dispositivo para a deteção de obstáculos baseado em visão stereo

e num algoritmo de estimação do plano do chão de forma a ajudar pessoas com dificuldades visuais a atingir um maior nível de independência [4].

Esta solução utiliza uma grelha de coordenadas polares para mapear um cenário. É desenhado e dividido em 3 zonas um semicírculo com um raio de 4,5 metros nessa grelha. As 3 zonas definem os níveis de aviso fornecidos ao utilizador.

De forma a determinar o nível do chão, é usado um algoritmo que analisa alguns pontos da grelha com base na sua altura, ignorando os pixéis que estão acima de uma determinada altura. A utilização deste algoritmo permite ignorar os pixéis que constituem o chão, fazendo com que não sejam considerados como possíveis obstáculos, e também estimar a altura de todos os pixéis no cenário, permitindo desta forma ignorar obstáculos que se encontrem a uma altura demasiado elevada.

Para fazer a deteção de obstáculos, são analisados todos os *frames* da sequência de vídeo, varrendo todos os pixéis da imagem que se encontrem dentro do limite da grelha e que não façam parte do chão. Estes pixéis são projetados na grelha e são determinadas as coordenadas polares dos mesmos. Por omissão, todos os pontos que tenham uma altura superior a 1,9 m são ignorados por ser considerado que estes não afetam pessoas com dificuldades visuais. São considerados como obstáculos as zonas que contenham pelo menos 500 pontos numa imagem com uma resolução de 640×480.

Para alertar o utilizador da posição e proximidade de obstáculos são usados sons curtos (*beeps*). Estes sons possuem uma frequência para objetos frontais e outra para objetos laterais. Estes sons são enviados para o auscultador do lado em que se encontra o obstáculo e caso o obstáculo se encontre no centro, são reproduzidos sons nos 2 auscultadores. A proximidade de um obstáculo é indicada pela repetição do som, quaias mais próximo o obstáculo mais vezes o som é tocado. O sistema apenas alerta para o obstáculo mais próximo, sendo que no caso de dois obstáculos laterais que se encontrem à mesma distância do utilizador, são reproduzidos os sons nos dois auriculares.

2.2.4 Seeing the Movement through Sound

Em 2014 uma equipa de investigação da Pontifícia Universidade Católica do Rio Grande do Sul no Brasil propôs um sistema que permitia que pessoas com dificuldades visuais

determinassem a trajetória de um objeto. O trabalho desta equipa propõe um modelo de sonorização que recebe as coordenadas 2D da localização de um objeto e gera um som com base nessas coordenadas. Este trabalho tinha principal objetivo o uso de som para descrever a trajetória descrita por um objeto a pessoas com dificuldades visuais.[5]

Com o recurso a uma *webcam* e ao *software CVMouse*, a posição de um objeto é convertida numa posição do rato no ecrã de um computador. Nesta posição é capturada a referência (coordenadas x e y) ao longo do tempo resultando assim num conjunto de coordenadas representativas do movimento do objeto. O objeto a ser seguido é determinado nos primeiros segundos da aplicação através da sua cor. Isto significa que não podem existir outros objetos com a mesma cor.

Para cada *pixel* (coordenada x , y) na imagem capturada pela câmara, é calculada a diferença absoluta sobre os valores da mediana. Os pixéis que são definidos como fundo são aqueles que apresentam uma diferença superior a um valor predefinido. No fim é gerada uma imagem com os pixéis que representam o movimento do objeto em destaque.

Uma vez determinada a imagem representativa do movimento do objeto, as coordenadas dos pixéis que descrevem o movimento são colocadas num eixo de coordenadas cartesianas. Cada ponto é convertido numa nota da escala musical (coordenada x) na respetiva octava (coordenada y).

2.2.5 Seeing the Sound

Em 2015 uma equipa de investigação da italiana publicou um artigo que descrevia um sistema que se propunha a produzir um mapa acústico sobreposto a uma imagem ótica com o objetivo de localizar objetos com base no som emitido por estes.[6]

Este sistema utiliza múltiplos microfones omnidirecionais. Cada um destes microfones capta os sinais sonoros proveniente de uma localização específica e utiliza esses sinais para calcular a intensidade de cada pixel. Sinais provenientes de outras localizações não podem ser somados de forma coerente e, por isso, serão desprezados.

Esta solução apresenta uma proposta de uma câmara acústico-ótica. Esta câmara possui uma dimensão de $0,45 \times 0,45$ metros e é composta por 128 microfones, uma câmara de vídeo

colocada no centro do dispositivo. A imagem captada pela câmara de vídeo é sobreposta à imagem acústica sobre a forma de uma imagem codificada por cores que representam a intensidade de cada pixel, sendo as cores mais frias usada para representar a menor intensidade dos pixéis.

2.2.6 Scanner 3D de imagens parciais

Em 2015 foi publicada uma patente que descreve um *scanner* 3D composto por duas câmaras. Estas câmaras são usadas para capturar o padrão formado quando uma fonte de luz incide sobre um objeto. O objeto 3D final é formado pela sobreposição das capturas realizadas.[7]

As câmaras utilizadas fazem a captação de imagens 2D da reflexão da fonte de luz no objeto. Tanto as câmaras como a fonte de luz devem estar fixas relativamente umas às outras e montadas na cabeça do scanner. De forma a efetuar o varrimento do padrão de luz no objeto alvo, a cabeça do scanner é colocada num eixo mecânico. Uma vez que esta solução necessita que a posição da cabeça do scanner seja determinada com elevada precisão, os movimentos ao longo do eixo mecânico devem ser predeterminados.

As imagens captadas pelas câmaras apenas precisam de ser parciais, uma vez que os campos de visão das câmaras sobrepõem-se parcialmente. Isto permite diminuir o tempo necessário para realizar a digitalização do objeto. Os sensores das câmaras estão sincronizados de forma a capturar a reflexão do padrão de luz em simultâneo, permitindo gerar uma imagem semelhante a uma imagem capturada por uma câmara capaz de cobrir todo o campo de visão.

O padrão de luz pode ser gerado por um laser, LED ou fonte de luz branca, desde que estes emitam um padrão com uma largura predefinida. Este padrão pode ser visível à luz, infravermelhos ou ultravioleta.

2.3 Microsoft Kinect

A *Kinect* é um dispositivo de deteção de movimentos produzido pela *Microsoft*. Produzido inicialmente em novembro de 2010 para as consolas da *Microsoft*, sendo o suporte para computador adicionado mais tarde (Junho de 2011), este dispositivo permite a interação do utilizador com a consola/computador através de gestos ou voz.[8]

A primeira versão deste dispositivo foi lançada em novembro de 2010 para a *Xbox 360*, tendo o *kit* de desenvolvimento para computador sido lançado em junho de 2011. Este SDK permitia o desenvolvimento de aplicações em C++, C# e Visual Basic. A segunda versão do dispositivo para computador (disponível para a *Xbox One* desde 2013) e SDK viriam a ser lançados em 2014.

A *Kinect* para a *Xbox 360* consiste num projetor de infravermelhos e uma câmara que geram uma grelha a partir da qual se pode determinar a localização de um objeto num espaço tridimensional. Este dispositivo possui uma câmara RGB, um sensor de profundidade e um microfone, permitindo assim fazer a captura em 3D dos movimentos do corpo, reconhecimento de voz e facial.[8]

Em novembro de 2013 foi lançada a segunda versão da *Microsoft Kinect* juntamente com a *Xbox One*. Esta nova iteração apresenta uma maior precisão e pode ser usada em ambientes mais escuros. Possui um campo de visão 60% maior e consegue detetar 6 pessoas ao mesmo tempo. A câmara captura vídeos em 1080p que, para além de melhorar a qualidade da captura, disponibiliza um *input* de informação estável [9]. A versão para computador desta nova versão apenas seria lançada em outubro de 2014, juntamente com a segunda versão do SDK e um adaptador para ligar o dispositivo à corrente elétrica.

Devido às capacidades de captura em 3D deste dispositivo e da qualidade do seu sensor, a *Microsoft Kinect* pode ser usada como um *scanner* 3D. Para tal é necessário um computador com uma placa gráfica capaz de obter pelo menos 20 *frames* por segundo. Devem ser evitadas sombras e reflexos durante a digitalização. A distância entre o objeto que se pretende digitalizar e outros objetos deve ser o maior possível, de forma a minimizar o ruído captado. A *Kinect* deve ser colocada a aproximadamente um braço de distância do objeto. A digitalização pode ser feita rodando o objeto ou movendo a *Kinect* à volta do objeto [10].

2.4 LabVIEW

LabVIEW (Laboratory Virtual Instrument Engineering Workbench) é uma plataforma de desenho de sistemas e ambiente de desenvolvimento da *National Instruments*, tendo a sua primeira versão sido lançada em 1986. Esta plataforma é caracterizada por usar uma linguagem de programação gráfica e pela facilidade em integrar diversos dispositivos [11].

O *LabVIEW* possui um elevado número de bibliotecas com diversas funções matemáticas, aquisição de dados, geração e conversão de sinais, entre outros. Possui ainda um módulo que permite programação baseada em texto.

Devido à longevidade e popularidade desta plataforma e da possibilidade de os utilizadores estenderem as suas funções, existe também um elevado número de *add-ons* criados pela comunidade para além das bibliotecas disponibilizadas nativamente.

2.5 Formato STL

O formato STL (*StereoLithography*) descrevem a geometria da superfície de um objeto tridimensional sem a representação cor ou textura. Este formato é usado principalmente para impressão 3D.

Os ficheiros STL descrevem um conjunto de triângulos que formam a superfície de um objeto. Para cada um destes triângulos são armazenados os valores das coordenadas de cada vértice, bem como os componentes de um vetor unitário que deve ser colocado a apontar para o lado de fora do objeto [12].

Este formato deve obedecer a quatro regras: a regra do vértice, a regra da orientação, a regra do octante positivo e a regra da ordenação dos triângulos. Estas regras indicam que cada triângulo deve partilhar dois dos seus vértices com os triângulos adjacentes, que a orientação deve ser especificada de duas maneiras (direção do vetor unitário e ordenação dos vértices na direção contrária aos ponteiros do relógio), as coordenadas dos vértices devem ser todas positivas (permite poupar espaço) e os triângulos devem ser ordenados segundo a coordenada z [12].

Os ficheiros deste formato podem ser binários ou ASCII, sendo o formato binário mais utilizado, uma vez que é mais compacto.

3 Metodologia

Neste capítulo será apresentada a metodologia adotada na elaboração deste trabalho, começando por introduzir o tema do trabalho e em que aspetos é que o mesmo constitui uma questão de investigação, à qual esta dissertação procura formas de dar resposta.

Definida a questão de investigação é necessário descrever como decorreu o processo de investigação desta dissertação, passando pelas bases que o fundamentaram até ao final do processo. De seguida, foi necessário identificar quais as ferramentas que foram seleccionadas e utilizadas para conceber esta solução e, por fim, especificar os passos que levaram à solução final.

3.1 Problema

Como referido anteriormente, as pessoas que têm graves dificuldade de visão possuem dificuldade em detetar a existência de obstáculos ou em reconhecer a forma de um objeto sem recorrer ao contato físico. As soluções existentes atualmente não são capazes de fornecer aos utilizadores informação sobre a forma completa de um objeto.

Assim sendo, o desafio abordado neste trabalho passa por apresentar um protótipo de um sistema capaz de gerar uma representação tridimensional de um objeto ou cenário tridimensional e convertê-lo em som.

3.2 Processo de Investigação

Na etapa inicial de elaboração da dissertação foi necessário definir o método de investigação mais adequado ao âmbito deste trabalho. Definiram-se quais seriam as principais etapas a realizar desde a revisão da literatura até ao final.

Inicialmente foi necessário elaborar uma base sólida de informação que permitisse corroborar e proporcionar o enquadramento necessário para o leitor sobre o tema do trabalho.

Iniciou-se a revisão da literatura com a investigação de temas sobre como o som é percebido

pelos humanos e que soluções já existiam para a conversão de imagem em som.

A etapa seguinte foi escolher quais as ferramentas a utilizar para o desenvolvimento do protótipo. Devido ao seu baixo custo, inicialmente pensou-se em usar um *Raspberry Pi* e o seu módulo de câmara para fazer a captação dos objetos e a sua conversão em som. Devido às limitações do *Raspberry Pi*, optou-se por implementar o sistema num computador normal, recorrendo para isso ao *LabVIEW* para fazer o processamento dos objetos 3D e a uma *Microsoft Kinect* para capturar os objetos. Inicialmente pensou-se em usar apenas um único par de auscultadores para realizar a reprodução do som, mas, numa tentativa de aumentar a precisão da solução, optou-se por utilizar 3 pares de auscultadores. Esta escolha será analisada na secção seguinte.

Segue-se a fase de implementação e desenvolvimento de código. Esta foi uma fase bastante exaustiva e morosa. Esta fase teve como principais objetivos a captação de objetos tridimensionais e a sua conversão em som. Esta fase exigiu diversas calibrações no processo de geração dos sinais sonoros de forma a obter sons diferentes para cada auscultador.

A fase de análise de resultados permitiu determinar a viabilidade do protótipo através dos resultados obtidos nos testes realizados.

No decorrer de todas estas fases foram também elaborados uma patente e dois artigos.

3.3 Seleção das Ferramentas a usar

Como mencionado anteriormente, nesta fase realizou-se uma pesquisa de forma a determinar a melhor forma de implementar o protótipo. Nesta fase foi necessário determinar quais as ferramentas a utilizar para a captação dos objetos, qual o formato em que seriam guardados os objetos, qual a ferramenta para fazer a conversão do objeto em som e como iria ser reproduzido esse som.

Uma vez que a utilização do módulo de câmara do *Raspberry Pi* introduzia a necessidade de captar várias imagens para criar um objeto 3D e devido às próprias limitações técnicas do *Raspberry Pi*, a escolha para a captação dos objetos recaiu sobre uma *Microsoft Kinect* e foi escolhida a linguagem *LabVIEW* para desenvolver a aplicação responsável por fazer a

conversão dos objetos em som.

A escolha da Microsoft Kinect deveu-se à facilidade da sua utilização para realizar digitalizações 3D, à qualidade do seu sensor, à possibilidade de gerar e guardar os objetos no formato pretendido. A linguagem *LabVIEW* foi escolhida por oferecer uma abordagem de programação gráfica, o que permitiu desenvolver a aplicação rapidamente.

Para o formato dos objetos 3D foi escolhido o formato STL. O formato STL descreve a geometria de superfície de um objeto em três dimensões, sem representar cor, textura ou outros atributos.

Para a reprodução do som foram criados uns *headphones* específicos. Estes *headphones* são compostos por 3 pares de auscultadores que encaixam numa estrutura impressa por uma impressora 3D. Esta solução fez com que não fosse necessário gastar dinheiro nuns *headphones* com as características necessárias, permitindo assim reduzir custos.

3.4 Desenvolvimento do Protótipo

Após selecionadas as ferramentas a utilizar, o passo seguinte passou pela implementação da aplicação responsável pela conversão dos objetos 3D em sinais sonoros.

O primeiro passo passou por implementar uma forma de alterar um sinal sonoro para que este fosse reproduzido de forma diferente para cada auscultador. Numa primeira etapa apenas se procurou fazer a distinção entre esquerda e direita. Após concluída esta etapa, procurou-se modificar o som para os seis auscultadores. Para esta etapa foi usado um sinal sonoro gerado artificialmente.

De seguida passou-se para a conversão de objetos criados artificialmente em som. Este passo permitiu simular a decomposição de um objeto real em camadas e gerar o som correspondente a cada ponto da camada. Desta forma foi possível detetar e corrigir um primeiro conjunto de erros sem a necessidade de usar objetos reais, uma vez que os objetos artificiais podiam ser facilmente alterados para os testes.

O passo seguinte passou pela utilização de objetos reais na geração do som. Neste passo foi

necessário fazer a captura de um objeto real e fazer a sua decomposição em camadas. Para cada camada foi necessário encontrar os pontos que formavam o seu contorno e gerar um som correspondente para esse ponto para cada auscultador.

4 Solução Proposta

Neste capítulo é apresentada a solução proposta para o problema apresentado nesta dissertação. Começa-se por fazer a descrição teórica de todo o processo de digitalização e conversão de objetos, seguindo-se a identificação o *hardware* utilizado e por fim é explicado o funcionamento do *software* desenvolvido.

4.1 Descrição teórica do processo

O processo começa com a utilização de um *scanner* 3D para realizar a digitalização de um objeto. Este objeto pode ser decomposto num sistema de coordenadas cartesianas tridimensional e representado em diversas camadas (planos paralelos ao eixo z) (Figura 1). O próximo passo é a utilização de imagens dos contornos do objeto em cada camada. Cada camada é mapeada num plano 2D, representando os contornos do objeto. Todos estes planos são representados por um conjunto de pontos que simulam um sinal sonoro num ambiente 3D.

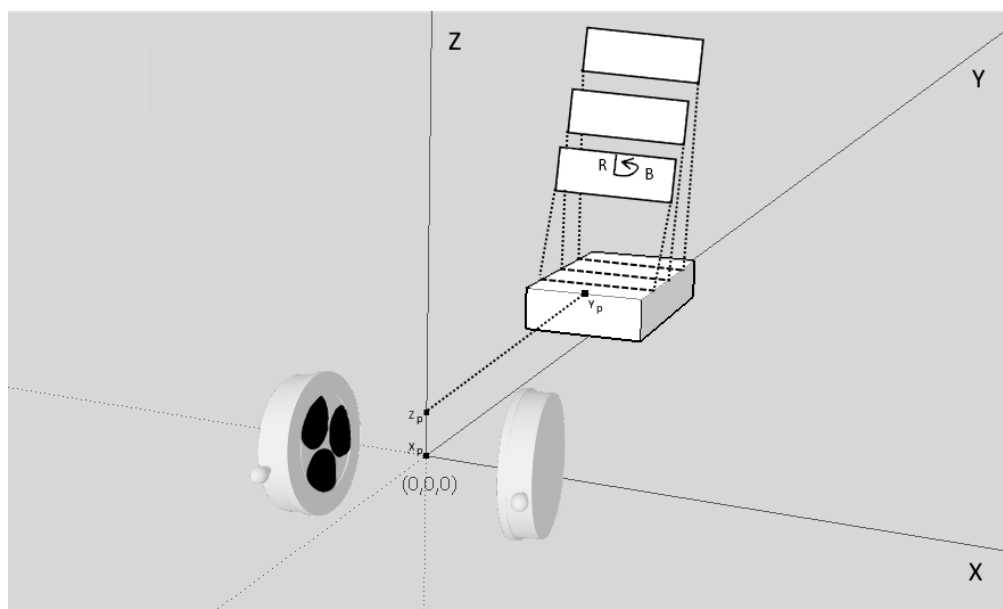


Figura 1 - Decomposição de um objeto 3D, por camada, no sistema Cartesiano

A localização dos pontos é representada por um conjunto de coordenadas polares 2D (raio = R e ângulo = B) centradas com uma linha horizontal que atravessa o centro do objeto tridimensional do espaço e o utilizador. Os pontos possuem frequências audíveis iguais sempre que os raios R sejam iguais, apesar dos ângulos B serem diferentes no intervalo 0-360°. Os pontos com um valor de R alto são representados por frequências audíveis baixas, enquanto os valores baixos de R são representados por frequências audíveis altas.

De forma a gerar o som de cada ponto para cada auscultador, é necessário determinar a distância a que este se encontra de cada auscultador, sendo utilizada a fórmula da distância euclidiana para esse efeito. Ou seja, a distância de cada ponto p para cada auscultador a pode ser calculada através da fórmula:

$$D = \sqrt{(x_p - x_a)^2 + (y_p - y_a)^2 + (z_p - z_a)^2}$$

Após determinadas as distâncias do ponto para cada auscultador, são gerados os sons para cada auscultador usando a fórmula do deslocamento:

$$y(t) = 256 * e^{-D} * \cos\left(\left(\frac{v}{D} * \frac{2\pi}{10^{-6} * D}\right) + (2\pi/(10^{-6} * D)) * \frac{t}{f}\right)$$

Na fórmula anterior é usada a frequência $f = 44,1$ kHz, considerou-se a velocidade de propagação do som no ar $v = 340$ m/s. A variável D corresponde à distância entre o ponto e o auscultador.

4.2 Identificação do *hardware* utilizado

De forma a realizar a conversão de um objeto ou cenário real num sinal sonoro é necessária uma forma de fazer a digitalização do objeto e uma forma de reproduzir o som.

Para fazer a digitalização de um objeto é necessário um scanner 3D capaz de disponibilizar a digitalização no formato STL. No caso deste protótipo é usada uma *Microsoft Kinect*.

Para fazer a reprodução do som são necessários uns *headphones* com 3 pares de auscultadores. Para este protótipo foram usados 3 pares de auscultadores *low cost* e foi usada uma impressora

3D para imprimir a estrutura dos *headphones*. Na figura 2 é possível ver um esquema de um dos lados dos *headphones* com os auscultadores.



Figura 2 - Representação de um dos lados dos headphones

Todo este sistema assenta num computador com o sistema operativo *Windows*. Na Figura 3 é possível ver parte do sistema utilizado.



Figura 3 - Headphones impressos

4.3 Modo de funcionamento do protótipo

A captura de um objeto ou cenário é feita através de uma *Microsoft Kinect*. De forma a obter

uma representação 3D com uma qualidade mínima para ser usada para gerar um som e de forma a não gerar um objeto com informação excessiva, a digitalização deve ser feita num curto espaço de tempo (aproximadamente 2 segundos). Pode ser feita uma digitalização mais demorada para gerar um objeto 3D mais detalhado, mas este incremento no detalhe apenas aumenta o tempo necessário para gerar o som e não apresenta diferenças significativas no sinal sonoro final.

Após concluída a digitalização do objeto, este é guardado no formato STL. Uma vez que este formato representa um objeto 3D num conjunto de triângulos e guarda as coordenadas (x, y, z) de cada um dos vértices de cada um dos triângulos, é possível fazer a divisão do objeto em camadas com base no valor da coordenada y .

Cada uma das camadas é representada num *array* bidimensional, em que cada posição do *array* que seja representativa de um ponto do contorno do objeto possui o valor da profundidade da camada e as restantes posições possuem o valor 0.

De seguida são calculados o ângulo e o raio de cada ponto de cada camada. Uma vez calculados todos os ângulos e raios, é gerado o som para os auscultadores ET (esquerdo traseiro), EC (esquerdo cima), EB (esquerdo baixo), DT (direito traseiro), DC (direito cima), DB (direita baixo). Na Figura 4 estão representados os 6 auscultadores e a distância D entre um ponto e cada um dos auscultadores. Para estes cálculos é necessário determinar as coordenadas de cada auscultador. Para tal são determinados os índices do centro do *array* e os auscultadores assumem os valores de x e z com base nesses índices, sendo o valor de y igual a 0 nos auscultadores de cima e baixo e igual a -10 nos auscultadores de trás. Uma vez terminado o processo de gerar os 6 sons, estes são reproduzidos pelo respetivo auscultador e é selecionada a próxima camada.

Todo o processo descrito pode ser visualizado no fluxograma presente na Figura 5.

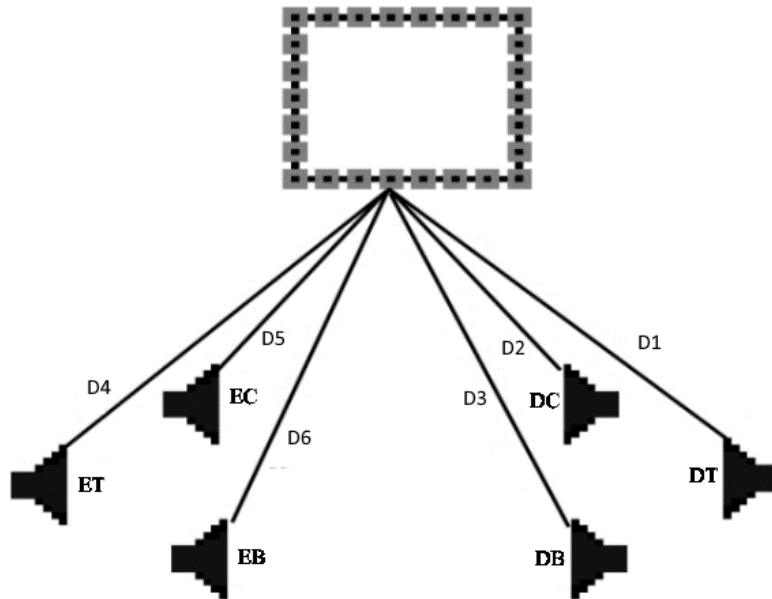


Figura 4 - Representação de uma camada, dos seus pontos e distâncias entre os auscultadores

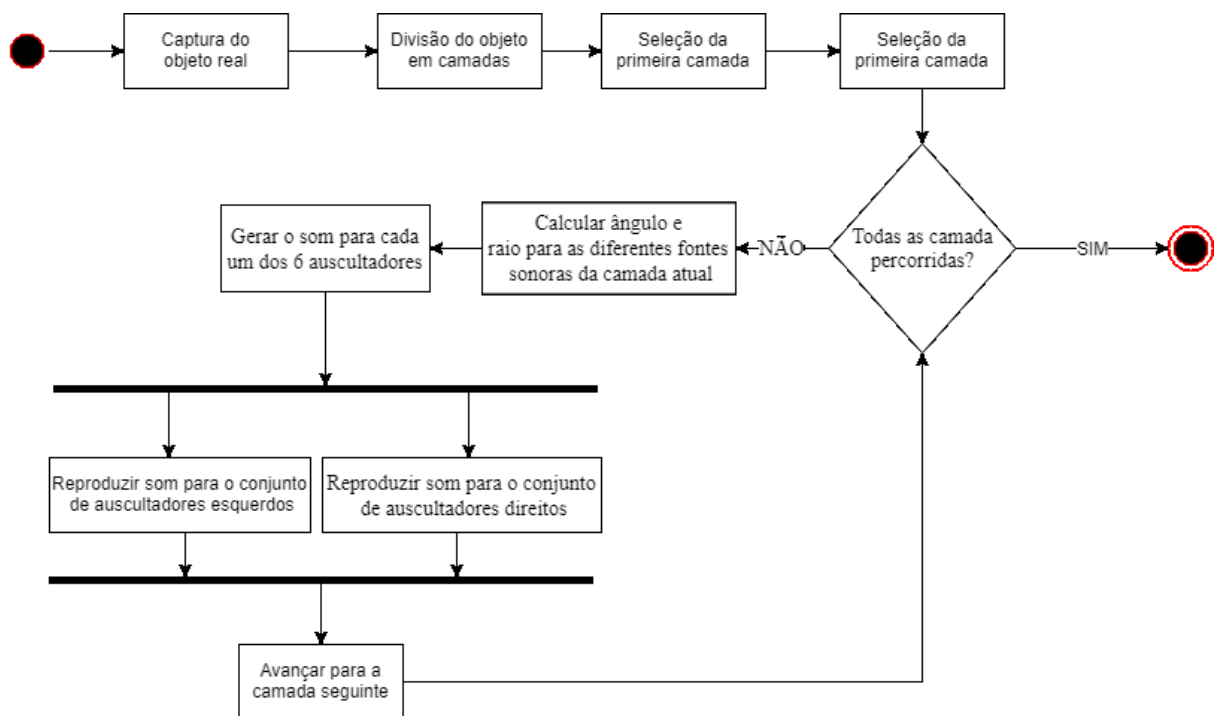


Figura 5 - Diagrama do funcionamento do protótipo

5 Discussão de Resultados

Nesta secção será apresentada a discussão dos resultados obtidos nos testes efetuados com o protótipo. Serão apresentados os resultados obtidos das conversões de diversos objetos 3D em som e os resultados dos testes efetuados com recurso a utilizadores reais.

5.1 Resultados da conversão de objetos em som

Numa primeira fase foram usados objetos gerados artificialmente de forma a que não fosse incluído ruído nos objetos. O objeto escolhido para esta fase foi uma esfera e o resultado da sua conversão em som pode ser observado na Figura 6.

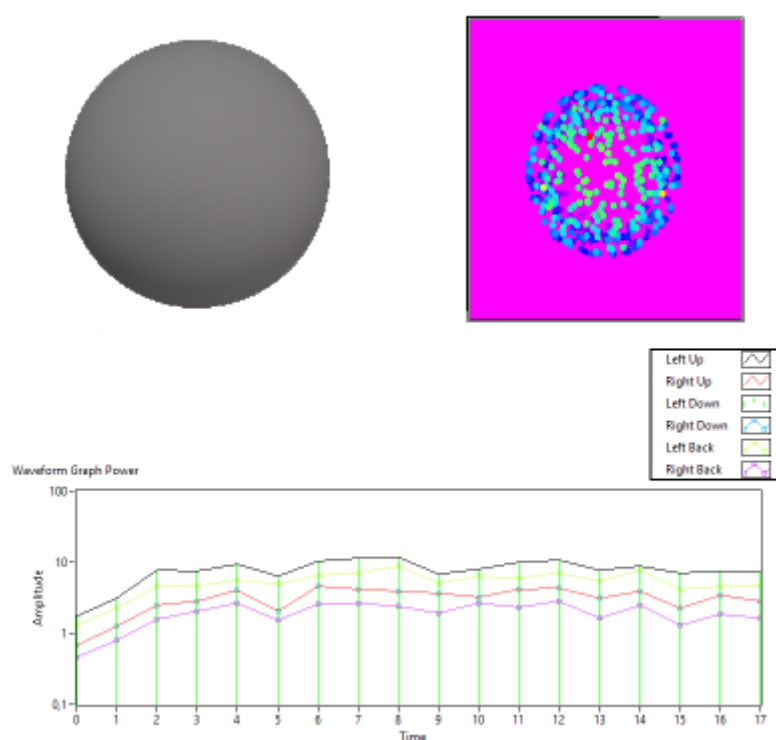


Figura 6 – Conversão de uma esfera

No lado superior esquerdo é possível observar o objeto 3D capturado, enquanto no lado superior direito é possível observar os pontos das diversas camadas sobrepostas. A cor de cada ponto representa uma frequência diferente. Na parte inferior da imagem é apresentado o gráfico da

onda sonora gerada para cada um dos 6 auscultadores.

A fase seguinte incidiu sobre a digitalização de objetos reais. Para a realização destes testes foram usados 3 objetos: uma porta fechada, uma porta semiaberta e uma pessoa com o braço estendido. Nas Figuras 7, 8 e 9 é possível observar o resultado da digitalização, a representação dos diversos pontos das várias camadas e o som gerado para cada um dos auscultadores.

A figura 6 mostra o resultado obtido para uma porta fechada. Uma vez que o resultado da digitalização apresentou uma ligeira inclinação do objeto, é possível observar também uma ligeira alteração nos pontos da porta. Pelo gráfico apresentado é possível verificar que o som é reproduzido durante um curto espaço de tempo. Isso devesse ao fato de se tratar de um objeto com pouca diferença de profundidade entre as camadas.

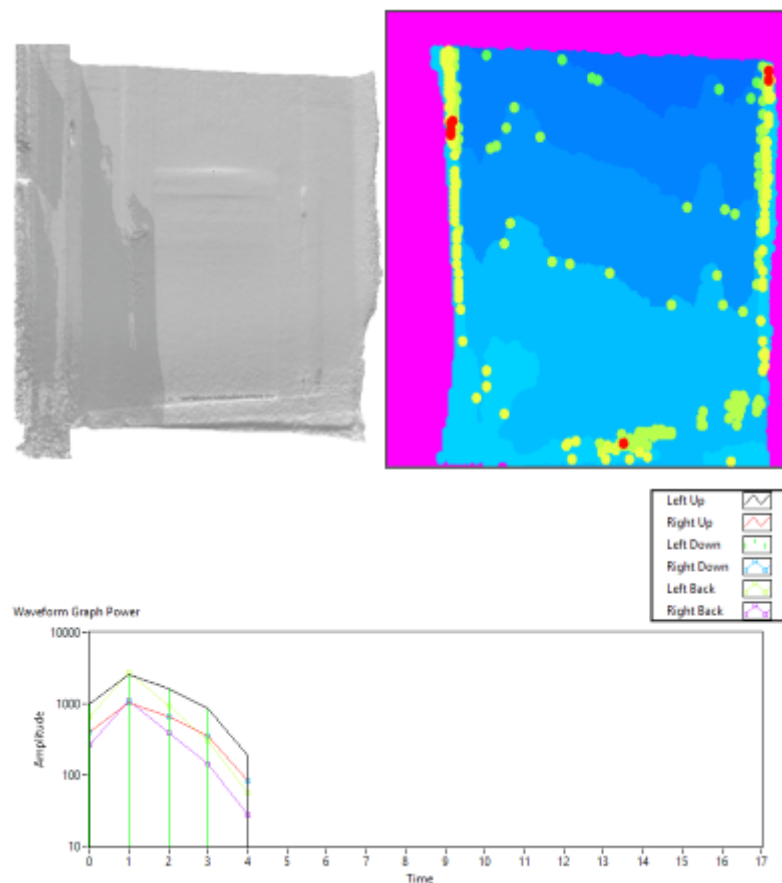


Figura 7 - Porta fechada

Na figura 8 é possível ver o resultado para uma porta semiaberta. Neste caso é possível identificar uma maior variação nos pontos de cada camada, ou seja, uma maior variação de profundidade. É também visível uma zona vazia, que representa a porção da porta que se

encontra aberta. É visível no gráfico que uma maior diferença de variação de profundidade faz com que o sinal sonoro seja reproduzido por mais tempo.

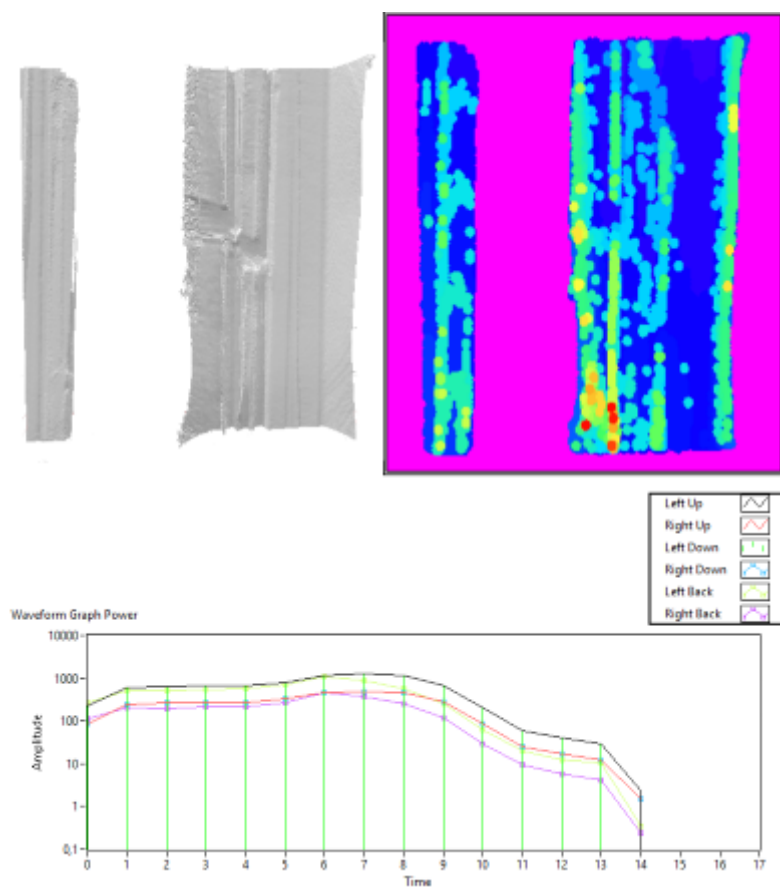


Figura 8 - Porta semiaberta

A figura 9 mostra o resultado de uma pessoa com a mão levantada. Neste caso é possível distinguir facilmente duas zonas com profundidades distinta: a zona do braço e a zona do corpo.

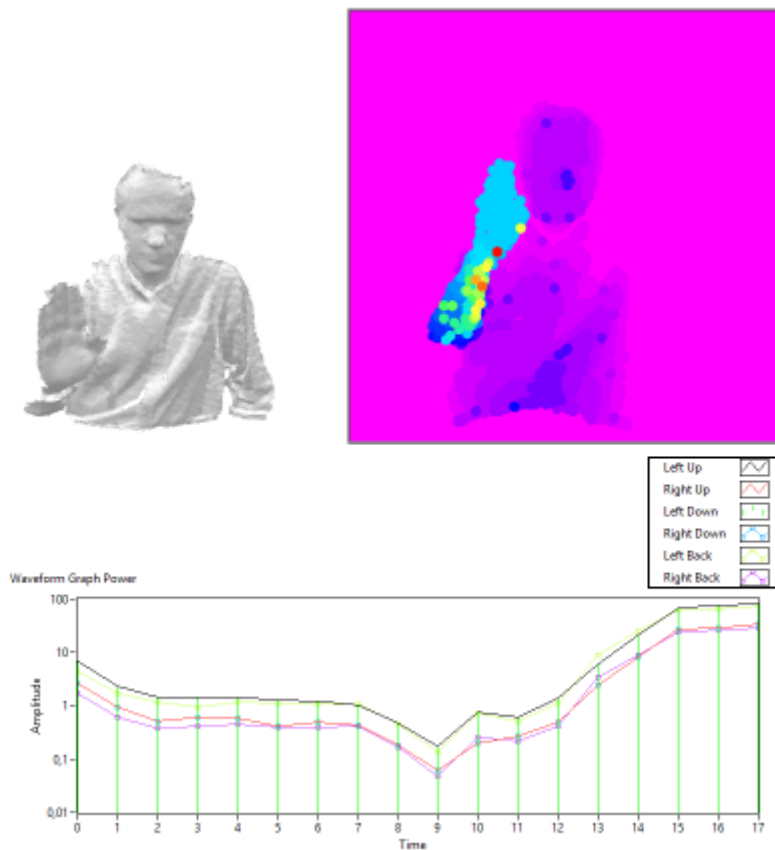


Figura 9 - Pessoa com o baço levantado

Comparando as figuras anteriores é possível concluir que a variação da profundidade do objeto influencia o tempo de reprodução do som, ou seja, quanto maior a profundidade, maior o tempo de reprodução.

5.2 Utilizadores reais

De forma a poder testar melhor o protótipo, foi utilizado um grupo de cinco pessoas. Há data da realização dos testes, destas cinco pessoas nenhuma possuía problemas visuais. Estes testes focaram-se apenas na parte de reprodução do som.

O teste consistiu em reproduzir o som de uma porta aberta cinco vezes e o som de uma porta semiaberta cinco vezes de forma aleatória. Os utilizadores sabiam a que objeto correspondia cada som. Após a reprodução dos dez sons, foram reproduzidos dez novos sons, cinco correspondentes a uma nova porta semiaberta e outros cinco correspondentes a uma nova porta aberta, de forma aleatória e foi pedido que indicassem a que objeto correspondia o som. Os resultados podem ser observados na seguinte tabela:

	<i>Pessoa 1</i>	<i>Pessoa 2</i>	<i>Pessoa 3</i>	<i>Pessoa 4</i>	<i>Pessoa 5</i>
<i>Objeto 1</i>	Errado	Errado	Certo	Certo	Certo
<i>Objeto 2</i>	Errado	Certo	Certo	Errado	Errado
<i>Objeto 3</i>	Certo	Errado	Errado	Errado	Certo
<i>Objeto 4</i>	Certo	Certo	Certo	Certo	Certo
<i>Objeto 5</i>	Certo	Certo	Certo	Certo	Certo
<i>Objeto 6</i>	Certo	Certo	Certo	Certo	Certo
<i>Objeto 7</i>	Certo	Certo	Certo	Certo	Certo
<i>Objeto 8</i>	Certo	Certo	Certo	Certo	Certo
<i>Objeto 9</i>	Certo	Certo	Certo	Certo	Certo
<i>Objeto 10</i>	Certo	Certo	Certo	Certo	Certo

Tabela 1 - Resultados testes utilizadores reais

Estes resultados mostram que o grupo conseguiu identificar corretamente a que objeto correspondia o som 42 vezes em 50 possíveis. Isto corresponde a uma taxa de respostas corretas de 84%. Também é possível verificar que as respostas erradas ocorreram nos primeiros sons reproduzidos.

6 Conclusão

Neste capítulo serão tecidas conclusões acerca do trabalho realizado, dos respectivos resultados, das limitações encontradas e dos objetivos alcançados. Finalmente apresentam-se os aspectos a ter em conta em trabalho futuro.

6.1 Síntese do Trabalho Realizado

Neste trabalho foi feito uma análise de como são gerados sons 3D e um levantamento de soluções que realizassem a conversão de imagens em som, com a intenção de desenvolver uma solução capaz de realizar a conversão de objetos reais em som com a melhor precisão possível.

Após realizadas esta análise e levantamento foi necessário determinar o *hardware* necessário para implementar a digitalização de objetos e para fazer a reprodução dos sons representativos dos objetos, e definir em que plataforma seria desenvolvida a aplicação da conversão desses mesmos objetos.

Após a identificação de todo o material necessário foi desenvolvida uma aplicação capaz de realizar a digitalização de diferentes objetos e posteriormente fazer a sua conversão em som. Uma vez concluída a aplicação, foram realizados os testes que permitiram comprovar a fiabilidade da mesma.

A solução desenvolvida foi ainda alvo de um pedido de patente (Anexo 1), bem com da escrita de dois artigos, um que viria a ser publicado na conferência “India's Digital Transformation” realizada na universidade PES na Índia (Anexo 2) e outro que foi aceite na conferência “IEEE Sponsored 4TH International Conference for Convergence in Technology (I2CT) 2018” em Mangalore na Índia, mas ainda não foi publicado (Anexo 3).

6.2 Resultados

Em comparação com outras soluções que usam imagens bidimensionais, a solução descrita nesta dissertação permite um melhor reconhecimento dos objetos. Em vez de serem usadas as intensidades de pixels numa imagem, a separação de um objeto tridimensional em diversas

camadas fornece aos utilizadores uma melhor perceção do aspeto do objeto.

Os resultados observados no capítulo Discussão de Resultados permitiram demonstrar numa primeira fase ser possível identificar um objeto através do som das suas camadas. Os testes com utilizadores reais demonstraram que com algum treino é possível identificar objetos semelhantes a objetos já conhecidos.

6.3 Limitações

A solução encontrada possui duas grandes limitações: a sua portabilidade e a robustez dos headphones.

A necessidade da *Microsoft Kinect* de ser ligada à corrente elétrica e os seus requisitos mínimos a nível de placa gráfica fazem com que sejam necessários uma tomada elétrica e um computador (mesmo que seja um portátil de pequenas dimensões), reduzindo assim a mobilidade do utilizador.

Relativamente aos *headphones*, apesar de as caixas onde estão embutidos os auscultadores não apresentarem fragilidades, as duas bandas que assentam na cabeça do utilizador e que ligam as caixas uma à outra são algo frágeis. Estas bandas apenas conseguem manter os *headphones* na cabeça do utilizador um curto espaço de tempo, uma vez que estas não suportam o peso das caixas durante muito tempo.

6.4 Trabalho futuro

Os próximos passos no desenvolvimento da solução apresentada passam por corrigir as limitações identificadas anteriormente. O primeiro passo será redesenhar os *headphones* para que as bandas que assentam na cabeça do utilizador sejam mais resistentes. de seguida é necessário exportar a aplicação desenvolvida para um dispositivo que ofereça uma maior portabilidade.

Futuramente também será necessário realizar testes em cenários mais complexos e com um maior número de pessoas.

Bibliografia

- [1] “Blindness and visual impairment.” [Online]. Available: <http://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>.
- [2] “Áudio binaural.” [Online]. Available: <https://hookeaudio.com/what-is-binaural-audio/>.
- [3] “vOICe.” [Online]. Available: <http://www.seeingwithsound.com/>.
- [4] A. Rodríguez, J. Javier Yebes, P. F. Alcantarilla, L. M. Bergasa, J. Almazán, and A. Cela, “Assisting the visually impaired: Obstacle detection and warning system by acoustic feedback,” *Sensors (Switzerland)*, vol. 12, no. 12, pp. 17476–17496, 2012.
- [5] N. Z. Salamon, J. C. S. J. Junior, and S. R. Musse, “Seeing the Movement through Sound : Giving Trajectory Information to Visually Impaired People,” pp. 12–17, 2014.
- [6] A. Zunino, M. Crocco, and S. Martelli, “Seeing the Sound: A New Multimodal Imaging Device for Computer Vision,” pp. 693–701, 2015.
- [7] “Patente scanner 3D de imagens parciais,” EP 2 839 238 B1.
- [8] “Project Natal.” [Online]. Available: <http://download.microsoft.com/download/A/4/A/A4A457B3-DF5D-4BF2-AD4E-963454BA0BCC/ProjectNatalFactSheetMay09.zip>.
- [9] “Microsoft Kinect V2.” [Online]. Available: <https://www.engadget.com/2013/05/21/microsofts-new-kinect-is-official/?guccounter=1>.
- [10] “3D scanning with Windows 10 & Kinect.” [Online]. Available: <https://www.createeducation.com/blog/3d-scanning-windows-10-kinect/>.
- [11] “LabVIEW.” [Online]. Available: <http://www.ni.com/labview/pt/>.
- [12] “STL File Format.” [Online]. Available: <https://all3dp.com/what-is-stl-file-format-extension-3d-printing>.

Apêndices

Anexos

Anexo 1 – Patente nacional n.º WO 2017/221177 A1

(12) PEDIDO INTERNACIONAL PUBLICADO SOB O TRATADO DE COOPERAÇÃO EM MATÉRIA DE PATENTES (PCT)

(19) Organização Mundial da
Propriedade Intelectual
Secretaria Internacional



(10) Número de Publicação Internacional
WO 2017/221177 A1

(43) Data de Publicação Internacional
28 de Dezembro de 2017 (28.12.2017) **WIPO | PCT**

(51) Classificação Internacional de Patentes:

G06F 3/16 (2006.01) G09B 21/00 (2006.01)
G01C 21/20 (2006.01) H04S 7/00 (2006.01)
G06F 3/0481 (2013.01)

(71) Requerente: INSTITUTO POLITÉCNICO DE LEIRIA
[PT/PT]; Ua General Norton De Matos, Apartado 4133,
2411-901 Leiria (PT).

(72) Inventores: DA SILVA PEREIRA, João; Rua Dos Sil-
vas, 226, 2425-353 Carvide (PT). LOURENÇO ALMEI-
DA, Nuno Miguel; Rua Dos Serviços Sociais N.º6 Bloco C
1.º Direito, 2430-168 Marinha Grande (PT).

(21) Número do Pedido Internacional:

PCT/IB2017/053707

(22) Data do Depósito Internacional:

21 de Junho de 2017 (21.06.2017)

(25) Língua de Depósito Internacional:

Português

(26) Língua de Publicação:

Português

(30) Dados Relativos à Prioridade:

109485 23 de Junho de 2016 (23.06.2016) PT

(74) Mandatário: PEREIRA DA CRUZ, João; Rua Vitor Cor-
don, 14, 1249-103 Lisboa (PT).

(81) Estados Designados (sem indicação contrária, para todos
os tipos de proteção nacional existentes): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,

(54) Title: METHOD AND APPARATUS FOR CREATING A THREE-DIMENSIONAL SCENARIO

(54) Título: MÉTODO E APARELHO DE CRIAÇÃO DE UM CENÁRIO TRIDIMENSIONAL

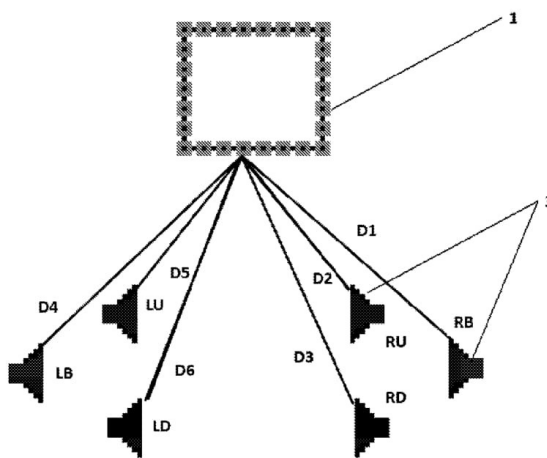


Figura 3

(57) Abstract: The present invention belongs to the field of virtual reality electronic equipment. The object of the present invention is a method for creating a three-dimensional scene, which method comprises the following steps: (a) to make a distance from sound-emitting means (3) - the distance from a virtual sound source - correspond to a sound with at least one frequency, a unique correspondence existing between a distance from a virtual sound source and a sound; and (b) emitting the sound using the sound-emitting means (3). The method thus allows a user to determine from the relationship between distance and frequency, without using the sense of sight, that there is a point in space relative to the user, and to measure the distance at which the point is located. In one embodiment, n virtual sound sources, each having its own frequency or associated frequencies, can be distributed on n front planes (1) to a user. The present invention also comprises a corresponding apparatus.

(Continua na página seguinte)

KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Estados Designados** (*sem indicação contrária, para todos os tipos de proteção regional existentes*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasiático (AM, AZ, BY, KG, KZ, RU, TJ, TM), Europeu (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Publicado:

- com relatório de pesquisa internacional (Art. 21(3))
- antes da expiração do prazo para modificar as reivindicações e a republicar na eventualidade de receção de tais modificações (Regra 48.2(h))

(57) **Resumo:** A presente invenção enquadra-se na área dos equipamentos electrónicos de realidade virtual. É objecto da presente invenção um método de criação de um cenário tridimensional que compreende: a) corresponder uma distância a meios de emissão de som (3) - denominada distância a uma fonte sonora virtual - a um som com pelo menos uma frequência, existindo uma correspondência única entre uma distância a uma fonte sonora virtual e um som; e b) emitir o referido som por via de meios de emissão de som (3). O método proporciona assim, pela relação entre distância e frequência, que um utilizador determine - sem recurso à visão - que um ponto espacial existe em relação a si, bem como uma medida da distância a que se encontra. Num modo de realização, é possível distribuir n fontes sonoras virtuais, cada qual com a sua frequência ou frequências associadas, por n planos frontais (1) a um utilizador. A presente invenção compreende igualmente um aparelho correspondente.

- 1 -

DESCRIÇÃO**MÉTODO E APARELHO DE CRIAÇÃO DE UM CENÁRIO TRIDIMENSIONAL****5 CAMPO DA INVENÇÃO**

A presente invenção enquadra-se na área dos equipamentos electrónicos de realidade virtual, tendo o presente invento aplicação directa na interacção com o espaço circundante por parte de pessoas com dificuldades visuais ou falta de percepção tridimensional de um espaço sem iluminação.

10

ANTECEDENTES DA INVENÇÃO

A presente invenção encontra antecedentes mais próximos em equipamentos de realidade virtual adequados para indivíduos que sofram de dificuldades visuais.

15

O pedido de patente com número de publicação EP2839238 descreve um sistema com duas câmaras adequado para captar um padrão criado a partir de uma fonte de luz refletida num objecto. As imagens captadas por cada câmara são sobrepostas de forma a criar um modelo tridimensional de um objecto detectado.

20

O artigo "Sight for Sore Ears" da revista IEEE Spectrum divulga um sistema chamado por vOICe, que inclui um dispositivo que funciona transformando imagens de uma câmara em imagens sonoras complexas, que são depois transmitidas ao utilizador através de auscultadores. O sistema ali descrito pode ser entendido como o estado da técnica mais próximo para a presente invenção, apresentando inúmeras limitações que

25

o presente invento vem resolver.

Concretamente, o sistema descrito no referido artigo apenas traduz uma imagem adquirida por uma única câmara num som, indicando qual a posição de um *pixel* na imagem e qual a cor, em tons de cinza, desse *pixel*. Trata-se portanto de uma solução

- 2 -

extremamente limitada, com utilidade bastante reduzida para um utilizador com dificuldades visuais que pretenda reconhecer e deslocar-se por um espaço circundante.

O objecto da presente invenção vem não só proporcionar uma solução
5 para esse problema como, por se tratar de uma solução mais capaz, inclui variados modos vantajosos que decorrem das capacidades melhoradas desta invenção.

SUMÁRIO DA INVENÇÃO

É assim objecto da presente invenção um método de criação de um cenário
10 tridimensional que compreende os seguintes passos:

- a) corresponder um valor que representa uma distância a meios de emissão de som (3), denominado distância a uma fonte sonora virtual, a um som com pelo menos uma frequência, tratando-se de uma correspondência única entre uma distância a uma fonte sonora virtual e um som;
- 15 b) emitir o referido som por via de meios de emissão de som (3).

A presente invenção proporciona assim, pela relação entre distância e frequência, que um utilizador determine – sem recurso à visão – que um ponto espacial existe em relação a si, bem como uma medida da distância a que se encontra. Este
20 método simula a localização de objectos num espaço, objectos esses reais ou não.

Num modo vantajoso do método da presente invenção, para um conjunto de n fontes sonoras virtuais a distância a uma fonte sonora virtual é inversamente proporcional a qualquer frequência de som emitido. Este esquema
25 permite não só que um utilizador identifique qual a distância de determinado ponto existente no espaço circundante mas também que reconheça a distância de um ponto em relação a outros pontos existentes no espaço circundante, construindo assim uma concepção mental desse espaço.

- 3 -

O método até aqui descrito é passível de ser implementado para cenários tridimensionais criados computacionalmente, para fins de realidade virtual, mas principalmente – como foi já repetidamente indicado – para fins de reconhecimento e apresentação sonora de um cenário real circundante.

5

Assim, num outro modo vantajoso do método da presente invenção, combinável com qualquer um dos descritos, este compreende um passo prévio ao passo a), de obtenção de um cenário tridimensional, que consiste na aquisição de um cenário tridimensional real.

10

O método de obtenção de um cenário tridimensional compreende, entre outros passos, o de estimação de distância a pelo menos um ponto de um objecto detectado (2), que por sua vez compreende os seguintes passos:

- intersecção de um plano frontal (1) a meios de varrimento (4) do espaço circundante com pelo menos um objecto tridimensional detectado (2), e
- associação a esse plano de um número finito de fontes sonoras virtuais.

15

Este método permite adicionalmente seccionar o espaço circundante, concretamente o espaço frontal a meios de varrimento (4) do espaço e portanto à frente de um utilizador, sendo que cada plano conterá um número finito de fontes sonoras virtuais. Assim, é possível distribuir n fontes sonoras virtuais, cada qual com a sua frequência ou frequências associadas, por n planos frontais (1) a um utilizador.

20

Neste sentido, e num modo vantajoso do método da presente invenção, combinável com qualquer um dos anteriores, a emissão de sons é realizada correspondendo cada instante temporal de emissão a um plano frontal (1) com um número finito de fontes sonoras virtuais, de forma sequencial no tempo e na distância do plano frontal (1) aos meios de varrimento (4).

25

- 4 -

Assim, trata-se de uma forma simples e claramente perceptível para um utilizador compreender o espaço circundante sem recurso à visão, através da associação de fontes sonoras virtuais em planos frontais (1), aos quais corresponde uma grandeza física – a sua distância a meios de emissão de som (3), que estarão solidários com um
5 utilizador –, representando dessa forma o espaço. A emissão de sons correspondentes a fontes sonoras virtuais agrupadas num plano mais distante dos meios de varrimento (4) ocorrerá antes da emissão de sons de um plano mais próximo. Assim, o utilizador apercebe-se da forma dos objectos com base em planos seccionados sequenciais, que poderão conter uma pluralidade de fontes sonoras virtuais a diferentes distâncias de si.

10

Numa outra configuração vantajosa do método da presente invenção, combinável com qualquer uma das anteriores, os meios de emissão de som (3) são pelo menos dois emissores de som, estando cada um dos emissores de som disposto de tal forma que um utilizador identifique a posição relativa do dito emissor em relação a si, e
15 configurados de tal forma que cada emissor emite um som de uma fonte sonora virtual de acordo com a posição relativa dessa fonte sonora ao utilizador.

Tal configuração garante mais um nível de compreensão do espaço circundante por parte de um utilizador, pois possibilita que este identifique se uma fonte
20 sonora virtual está em determinada posição em relação a si de acordo com os emissores de som que emitem som em determinado instante.

Faz igualmente parte da presente invenção um aparelho de criação de um cenário tridimensional que compreende meios de emissão de som (3) configurados
25 para, sob um valor que representa uma distância a um utilizador – denominada distância a uma fonte sonora virtual –, emitirem um som com pelo menos uma frequência, sendo esse um som único correspondente a essa fonte sonora virtual.

- 5 -

Este aparelho materializa num objecto físico as vantagens do método já descrito, possibilitando que um utilizador determine – sem recurso à visão – que um ponto espacial existe em relação a si.

5 Preferencialmente, este aparelho está configurado para implementar o método acima descrito, nos diferentes níveis de detalhe descritos, as suas diferentes configurações.

Num modo vantajoso do aparelho da presente invenção, os meios de
10 emissão de som (3) compreendem pelo menos três emissores de som, estando cada um dos três emissores de som disposto de tal forma que um utilizador identifique a posição relativa do dito emissor em relação a si.

Tal configuração materializa as vantagens já anteriormente descritas para
15 o método da presente invenção, adicionando um nível de percepção do espaço ao utilizador.

DESCRIÇÃO DAS FIGURAS

O presente conjunto de figuras refere-se a modos de realização específicos do presente
20 invento, não pretendendo portanto limitá-lo, mas apenas melhor ilustrar esses modos de realização.

Figura 1 – Representação de um objecto e dos seus planos fontais de corte, onde R representa a distância de centro do plano até um ponto e B representa o ângulo. No
25 eixo X estão presentes três pares de altifalantes (3) juntamente com o par de meios de varrimento (4) de tipo ultrassónico e um par de meios de varrimento (4) com câmaras (5). O dispositivo da invenção é aqui representado por dois cubos que estão localizados sobre o eixo X e centrado por simetria na origem (0, 0, 0). A forma real desse dispositivo é semelhante a um par de auscultadores que contem dois ares de meios de varrimento
30 (4) 3D (funcionando por ultrassons e imagens) e três pares de altifalantes (3). O objeto

- 6 -

3D que se quer detetar por som é representado por uma forma retangular e os seus planos frontais (1) de corte estão paralelos ao plano XZ. Um dos pontos desse objeto está posicionado nas coordenadas (x_p, y_p, z_p) .

5 Figura 2 – Representação de 1 a N planos frontais (1) de corte de um objecto e dos diferentes pontos presentes em cada um dos planos. Os pontos serão utilizados para simular o posicionamento de uma fonte sonora virtual. São apresentados planos de um objecto rectangular com N cortes frontais paralelos ao plano XZ da figura 1.

10 Figura 3 – Representação da distância entre um dos pontos de um plano de corte e diversos altifalantes (3) – os meios de emissão de som (3). São apresentadas distâncias D1, D2, ..., D6 entre um ponto (de uma fonte sonora virtual) e 6 altifalantes (3). Para cada ponto P é calculada a distância euclidiana com cada altifalante (3) (Direita Trás (RB), Direita Cima (RU), Direita Baixo (RD), Esquerda Trás (LB), Esquerda Cima (LU), Esquerda

15 Baixo (LD)), num total de seis distâncias por ponto:

$$\bullet \quad D1 = \sqrt{(x_p - x_{RB})^2 + (y_p - y_{RB})^2 + (z_p - z_{RB})^2}$$

$$\bullet \quad D2 = \sqrt{(x_p - x_{RU})^2 + (y_p - y_{RU})^2 + (z_p - z_{RU})^2}$$

$$\bullet \quad D3 = \sqrt{(x_p - x_{RD})^2 + (y_p - y_{RD})^2 + (z_p - z_{RD})^2}$$

$$\bullet \quad D4 = \sqrt{(x_p - x_{LB})^2 + (y_p - y_{LB})^2 + (z_p - z_{LB})^2}$$

20 $\bullet \quad D5 = \sqrt{(x_p - x_{LU})^2 + (y_p - y_{LU})^2 + (z_p - z_{LU})^2}$

$$\bullet \quad D6 = \sqrt{(x_p - x_{LD})^2 + (y_p - y_{LD})^2 + (z_p - z_{LD})^2}$$

Esses cálculos são utilizados para simular o posicionamento de uma fonte sonora virtual cujo sinal deverá se propagar até chegar a cada um dos seis altifalantes (3), utilizando
25 um modelo ideal de propagação do som em função da distância de cada um dos receptores.

Figura 4 – Esquema representativo de um aparelho de acordo com a presente invenção,
30 compreendendo altifalantes (3), câmara (5) e radares ultrassónicos (4).

- 7 -

DESCRIÇÃO DETALHADA DA INVENÇÃO

Os principais modos vantajosos do objecto da presente invenção estão descritos na secção Sumário da Invenção, descrevendo-se de seguida particularidades decorrentes

5 desses modos vantajosos.

10 Numa configuração preferida do método da presente invenção, combinável com qualquer uma das anteriores, a emissão de n sons correspondentes a n fontes sonoras virtuais é realizada periodicamente. Tal configuração permite que um utilizador reconheça repetidamente o espaço circundante, através da emissão repetida de som correspondente às fontes sonoras virtuais que representam um espaço circundante. Adicionalmente, tal possibilita actualização dos sons que representam o espaço, por exemplo como consequência de movimento do utilizador.

15 Numa outra configuração do método da presente invenção, combinável com qualquer uma das anteriores, a aquisição de um cenário tridimensional real compreende os seguintes passos:

- varrimento de um espaço circundante a um utilizador através de meios de varrimento (4);
- 20 • detecção de objectos potencialmente presentes no espaço circundante;
- estimação de distância a pelo menos um ponto de um objecto;
- classificação da distância estimada no passo anterior como uma distância a uma fonte sonora virtual.

25 Assim, para um cenário real, uma distância a um ponto de um objecto detectado (2) é medida e é-lhe associada uma fonte sonora virtual, a qual tem pelo menos uma frequência associada. Tal ocorre por varrimento com base em meios adequados do espaço circundante, detecção de objectos que possam estar presentes nesses espaço e estimação de distância a pelo menos um ponto do objecto. Quanto mais

- 8 -

pontos forem utilizados para representar o espaço circundante, mais complexa e completa será essa representação.

5 Numa outra configuração do método da presente invenção, combinável com qualquer uma das anteriores, os meios de varrimento (4) são pelo menos dois, sendo que a detecção de objectos potencialmente presentes no espaço circundante compreende o cálculo da média dos sinais obtidos dos pelo menos dois meios de varrimento (4).

10 Tal possibilita uma melhor detecção de objectos, por via de ser usado um par de meios de varrimento (4). Para estimação de um só objecto, é realizada a média dos objectos estimados.

15 Numa outra configuração do método da presente invenção, os meios de varrimento (4) são ultrassónicos e/ou ópticos, sendo que a referida média é realizada para todos os objectos estimados por via dos diferentes meios de varrimento (4).

20 Numa outra configuração do método da presente invenção, combinável com qualquer uma das anteriores, qualquer frequência emitida se encontrar na gama do audível para um ser humano.

25 Numa outra configuração do método da presente invenção, combinável com qualquer uma das anteriores, é emitido um som caso seja detectada uma alteração num dos referidos planos frontais (1).

 Tal possibilita que o utilizador percepcione com maior clareza alterações no espaço circundante.

30 Num modo específico do aparelho da presente invenção, combinável com qualquer um dos anteriormente descritos, este compreende meios de varrimento

- 9 -

(4), preferencialmente configurados para implementar o método anteriormente descrito, em qualquer uma das suas configurações.

Num modo específico daquele imediatamente acima descrito, os meios
5 de varrimento (4) são ultrassónicos e/ou ópticos.

O aparelho da presente invenção compreende igualmente pelo menos um controlador – que inclui meios computacionais – para processamento de dados e interface e controlo de qualquer um dos restantes elementos.

10

Num modo específico daquele descrito imediatamente acima, os meios de emissão de som (3) consistirem em seis altifalantes (3) agrupados três a três, os meios de varrimento (4) consistirem em um par de sondas de ultrassons e um par de câmaras (5) sensíveis a radiação visível e infravermelha.

15

MODOS DE REALIZAÇÃO

Descrevem-se de seguida modos de realização do aparelho da presente invenção.

Esse aparelho tem uma estrutura externa com a forma de um par de
20 auscultadores de áudio. No interior dessa estrutura, está contido um par meios de varrimento (4) 3D ultrassónicos, um par de meios de varrimento (4) 3D com câmaras (5) sensíveis à luz visível e a infravermelhos e de 3 pares de altifalantes (3).

Podem ser usados meios de varrimento (4) tridimensional do tipo sonares
25 ou radares ultrassónicos, ou com câmaras (5) com uma determinada sensibilidade à luz e que possam determinar a deformação de um padrão sobre a superfície do objecto com base em diferentes orientações/posições das câmaras (5).

Os vários tipos de meios de varrimento (4) tridimensional são usados aos
30 pares para reforçar a precisão do cálculo da profundidade de um objecto ou cenário. O

- 10 -

par de meios ultrassônicos poderá ser constituído por emissores/receptores de ultrassons colocados sobre uma plataforma móvel que permita fazer periodicamente o varrimento do cenário de cima para baixo, da esquerda para a direita, e assim construir uma imagem 3D do mesmo.

5

Descreve-se de seguida um modo de realização do método da presente invenção.

Através da utilização de meios de varrimento (4) 3D é possível criar
10 objectos tridimensionais do espaço onde uma pessoa se encontra. A superfície desse objecto é virtualmente coberta por várias fontes sonoras virtuais. Para cada fonte sonora virtual localizada na superfície do objecto é calculada a distância entre o utilizador e essa fonte. Essas distâncias são calculadas de forma a simular as localizações das várias fontes sonoras provenientes do espaço tridimensional que chegam aos 3
15 pares de altifalantes (3), considerando que os sinais sonoros se propagam num meio de transmissão homogêneo, sem obstáculos, sem distorção e sem reflexões.

Esse objecto ou cenário espacial 3D é decomposto em várias camadas frontais paralelas que são acedidas/utilizadas periodicamente e sequencialmente,
20 sendo o varrimento periódico efetuado da camada mais afastada até à mais próxima. São utilizadas frequências audíveis diferentes para determinar cada um dos planos frontais (1), utilizados em cada momento. Cada uma dessas camadas é representada num plano 2D, onde estão localizadas curvas por onde passam os cortes dos planos frontais (1) com o objecto 3D. As curvas de cada plano frontal (1) são representadas por
25 uma quantidade limitada de pontos que são utilizados para simular a proveniência de uma fonte sonora num espaço tridimensional. Os pontos de localizações virtuais das fontes sonoras virtuais, de cada um dos planos, são representados por coordenadas polares 2D (raio = R e ângulo = B) centradas com uma linha horizontal que atravessa o centro do objecto tridimensional do espaço e o centro dos 3 pares de altifalantes (3). Os
30 pontos de localização virtuais das fontes sonoras têm frequências audíveis iguais sempre

- 11 -

que os raios R sejam iguais, apesar dos ângulos B poderem ser diferentes no intervalo [0°, 360°]. Os pontos virtuais com raios R elevados são representados por frequências audíveis baixas e os pontos com raios baixos são representados por frequências audíveis mais altas. O utilizador da invenção pode estimar o contorno do objecto com base na
5 audição de uma frequência audível proporcional ao raio R em cada plano. Esse processo repete-se periodicamente e rapidamente em cada plano frontal (1) com frequências diferentes.

Os 3 pares de altifalantes (3) efectuem a emissão de som com base na
10 simulação das diversas fontes sonoras virtuais espalhadas num espaço tridimensional que é varrido periodicamente de trás para frente e das extremidades para o centro em cada plano frontal (1) individual.

Os 3 pares de altifalantes (3) estão localizados próximo do sistema
15 auditivo de utilizador para que ele tenha a sensação de captação/audição de som envolvente proporcional a forma do objecto 3D ou do cenário tridimensional. Cada um dos pares de altifalantes (3) está localizado convenientemente para proporcionar ao utilizador a sensação da proveniência correta do som (em cima, em baixo e atrás, com espaçamento de alguns centímetros junto de cada orelha). Ou seja, esse som pode ser
20 personalizado com uma orientação/direção: "vem de cima ou de baixo", "vem da direita ou da esquerda" e "vem de frente ou de trás". O varrimento periódico de todos os planos frontais (1) proporciona ao utilizador sons distintos para cada tipo de forma tridimensional.

25 Após adquirida as várias imagens 3D provenientes dos vários meios de varrimento (4) é efetuado uma média das várias imagens 3D. Depois desse processo, o objecto 3D é decomposto em vários planos frontais (1) onde são desenhados as várias linhas de cortes.

- 12 -

Como será evidente a um perito na especialidade, a presente invenção não deverá estar limitada aos modos de realização descritos no presente documento, sendo possíveis diversas alterações que se mantêm no âmbito da presente invenção.

- 5 Evidentemente, os modos preferenciais acima apresentados são combináveis, nas diferentes formas possíveis, evitando-se aqui a repetição de todas essas combinações.

- 13 -

REIVINDICAÇÕES

1. Método de criação de um cenário tridimensional **caracterizado por** compreender os seguintes passos:

- 5 a) corresponder um valor que representa uma distância a meios de emissão de som (3) – denominado distância a uma fonte sonora virtual – a um som com pelo menos uma frequência, existindo uma correspondência única entre uma distância a uma fonte sonora virtual e um som;
- b) emitir o referido som por via de meios de emissão de som (3).

10

2. Método de acordo com a reivindicação anterior **caracterizado por**, para um conjunto de n fontes sonoras virtuais, a distância a uma fonte sonora virtual ser inversamente proporcional a qualquer frequência de som emitido.

15

3. Método de acordo com qualquer uma das reivindicações anteriores **caracterizado por** a emissão de n sons correspondentes a n fontes sonoras virtuais ser realizada periodicamente.

4. Método de acordo com qualquer uma das reivindicações anteriores **caracterizado por** os meios de emissão de som (3) serem pelo menos três emissores de som, estando cada um dos três emissores de som disposto de tal forma que um utilizador identifique a posição relativa do dito emissor em relação a si, e configurados de tal forma que cada emissor emita um som de uma fonte sonora virtual de acordo com a posição relativa dessa fonte sonora ao utilizador.

25

5. Método de acordo com qualquer uma das reivindicações anteriores **caracterizado por** compreender um passo prévio ao passo a), de obtenção de um cenário tridimensional, que consiste na aquisição de um cenário tridimensional real.

30

- 14 -

6. Método de acordo com a reivindicação anterior **caracterizado por** a aquisição de um cenário tridimensional real compreender os seguintes passos:

- varrimento de um espaço circundante a um utilizador através de meios de varrimento (4);
- 5 • detecção de objectos potencialmente presentes no espaço circundante;
- estimação de distância a pelo menos um ponto de um objecto;
- classificação da distância estimada no passo anterior como uma distância a uma fonte sonora virtual.

10 7. Método de acordo com a reivindicação anterior **caracterizado por** compreender um passo de estimação de distância a pelo menos um ponto de um objecto, que por sua vez compreende os seguintes passos:

- intersecção de um plano frontal (1) aos meios de varrimento (4) com pelo menos um objecto tridimensional detectado (2);
- 15 • associação a esse plano de um número finito de fontes sonoras virtuais.

8. Método de acordo com as reivindicações 2-6 **caracterizado por** a emissão de sons ser realizada correspondendo cada instante temporal de emissão a um plano frontal (1) com um número finito de fontes sonoras virtuais, de forma sequencial
20 no tempo e na distância do plano frontal (1) aos meios de varrimento (4).

9. Método de acordo com qualquer uma das reivindicações 5-7 **caracterizado por** os meios de varrimento (4) ultrassónicos e/ou ópticos serem pelo menos dois, sendo que a detecção de objectos potencialmente presentes no espaço
25 circundante compreende o cálculo da média dos sinais obtidos dos pelo menos dois meios de varrimento (4).

10. Método de acordo com qualquer uma das reivindicações anteriores **caracterizado por** qualquer frequência emitida se encontrar na gama do
30 audível para um ser humano.

- 15 -

11. Método de acordo com qualquer uma das reivindicações 7-10 **caracterizado por** emitir um som caso seja detectada uma alteração num dos referidos planos frontais (1).

5

12. Aparelho de criação de um cenário tridimensional compreendendo meios de emissão de som (3) **caracterizado por** os meios de emissão de som (3) estarem configurados para, sob um valor que representa uma distância a meios de emissão de som (3) – denominada distância a uma fonte sonora virtual –,
10 emitirem um som com pelo menos uma frequência, sendo esse um som único correspondente a essa fonte sonora virtual, preferencialmente configurado para implementar o método de qualquer uma das reivindicações 1-4.

13. Aparelho de acordo com a reivindicação anterior **caracterizado**
15 **por** os meios de emissão de som (3) compreenderem pelo menos três emissores de som, estando cada um dos três emissores de som disposto de tal forma que um utilizador identifique a posição relativa do dito emissor em relação a si.

14. Aparelho de acordo com qualquer uma das reivindicações 12-13
20 **caracterizado por** compreender meios de varrimento (4), preferencialmente configurados para implementar o método de qualquer uma das reivindicações 1-11, sendo esses meios de varrimento (4), preferencialmente, ultrassónicos e/ou ópticos.

15. Aparelho de acordo com a reivindicação anterior **caracterizado**
25 **por** os meios de emissão de som (3) consistirem em seis altifalantes (3) agrupados três a três, os meios de varrimento (4) consistirem em um par de sondas de ultrassons e um par de câmaras (5) sensíveis a radiação visível e infravermelha.

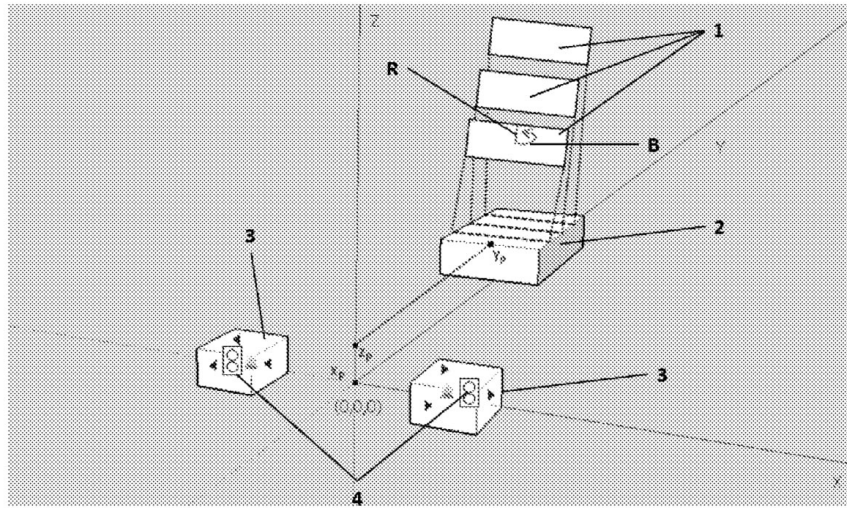


Figura 1

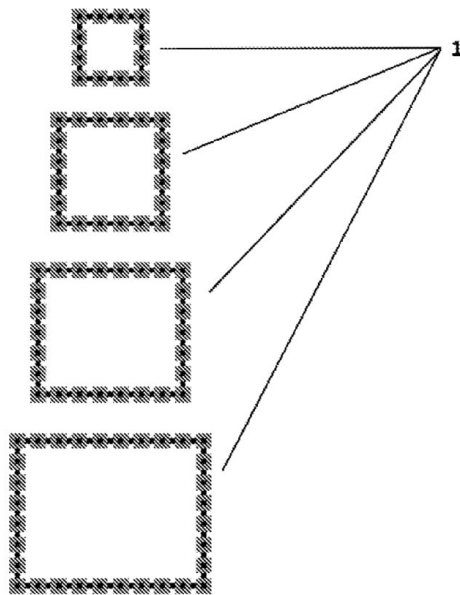


Figura 2

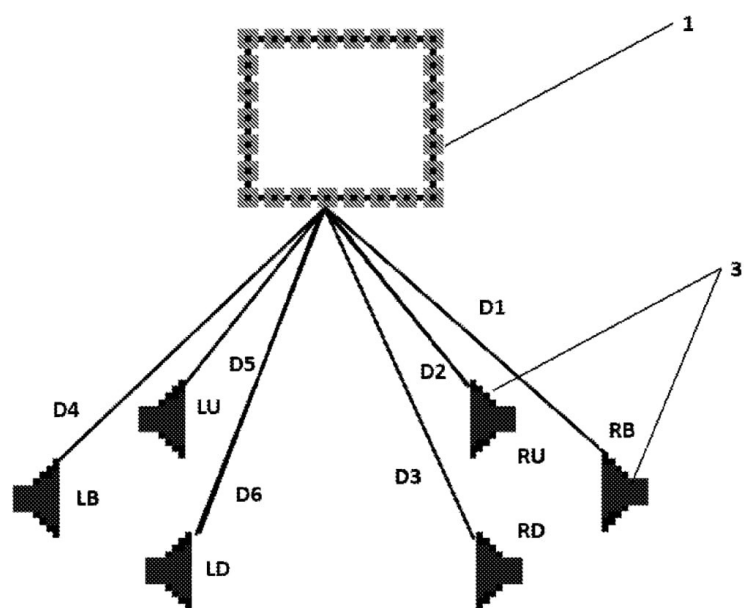


Figura 3

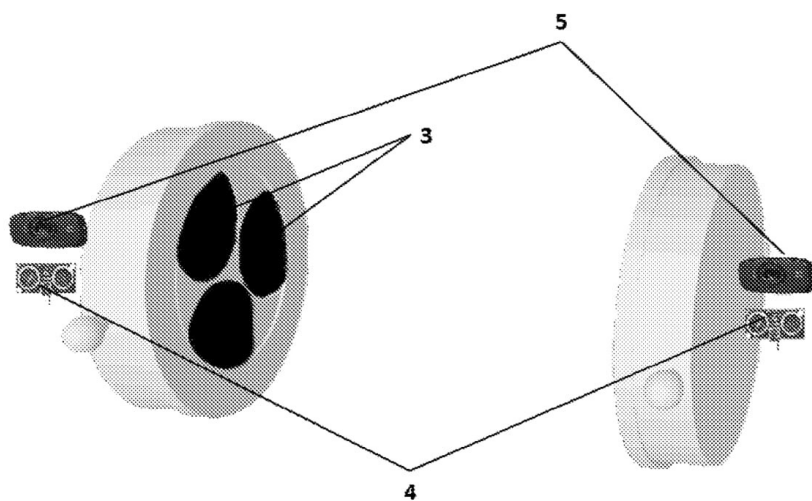


Figura 4

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2017/053707

A. CLASSIFICATION OF SUBJECT MATTER

INV. G06F3/16 G01C21/20 G06F3/0481 G09B21/00 H04S7/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F G01C G09B H04S H04R G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2015/198284 A1 (D AMICO ALESSIO MARIA [IT]; LOPPINI FABRIZIO [IT]) 30 December 2015 (2015-12-30) page 3, line 24 - page 5, line 25 page 9, line 18 - page 14, line 33 page 17, line 3 - page 27, line 23 figures 1,4	1-15
X	US 2011/188342 A1 (SPORER THOMAS [DE]) 4 August 2011 (2011-08-04) paragraphs [0004] - [0018] paragraphs [0036] - [0057] figures 2, 3A, 6	1-15

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

6 October 2017

Date of mailing of the international search report

17/10/2017

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Vilella, Josep

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/IB2017/053707

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2015198284 A1	30-12-2015	NONE	
US 2011188342 A1	04-08-2011	CN 101978424 A	16-02-2011
		EP 2255359 A1	01-12-2010
		JP 2011516830 A	26-05-2011
		KR 20100116223 A	29-10-2010
		US 2011188342 A1	04-08-2011
		WO 2009115299 A1	24-09-2009

Form PCT/ISA/210 (patent family annex) (April 2005)

RELATÓRIO DE PESQUISA INTERNACIONAL

Depósito internacional Nº

PCT/IB2017/053707

A. CLASSIFICAÇÃO DO OBJETO		
G06F3/16	G01C21/20	G06F3/0481 G09B21/00 H04S7/00
De acordo com a Classificação Internacional de Patentes (IPC) ou conforme a classificação nacional e IPC		
B. DOMÍNIOS ABRANGIDOS PELA PESQUISA		
Documentação mínima pesquisada (sistema de classificação seguido pelo símbolo da classificação)		
G06F G01C G09B H04S H04R G06K		
Documentação adicional pesquisada, além da mínima, na medida em que tais documentos estão incluídos nos domínios pesquisados		
Base de dados eletrônica consultada durante a pesquisa internacional (nome da base de dados e, se necessário, termos usados na pesquisa)		
EP0-Internal, WPI Data		
C. DOCUMENTOS CONSIDERADOS RELEVANTES		
Categoria*	Documentos citados, com indicação de partes relevantes, se apropriado	Relevante para as reivindicações Nº
X	WO 2015/198284 AI (D AMICO ALESSIO MARIA [IT]; LOPPINI FABRIZIO [IT]) 30 Dezembro 2015 (2015-12-30) página 3, linha 24 - página 5, linha 25 página 9, linha 18 - página 14, linha 33 página 17, linha 3 - página 27, linha 23 figuras 1,4	1-15
X	US 2011/188342 AI (SPORER THOMAS [DE]) 4 Agosto 2011 (2011-08-04) parágraphos [0004] - [0018] parágraphos [0036] - [0057] figuras 2, 3A, 6	1-15
<input type="checkbox"/> Documentos adicionais estão listados na continuação do Quadro C <input checked="" type="checkbox"/> Ver o anexo de família da patentes		
* Categorias especiais dos documentos citados: "A" documento que define o estado geral da técnica, mas não é considerado de particular relevância. "E" depósito ou patente anterior, mas publicada após ou na data do depósito internacional. "L" documento que pode lançar dúvida na(s) reivindicação(ões) de prioridade ou na qual é citado para determinar a data de outra citação ou por outra razão especial (como especificado). "O" documento referente a uma divulgação oral, uso, exibição ou por outros meios. "P" documento publicado antes do depósito internacional, porém posterior a data de prioridade reivindicada. "T" documento publicado depois da data de depósito internacional, ou de prioridade e que não conflita com o depósito, porém citado para entender o princípio ou teoria na qual se baseia a invenção. "X" documento de particular relevância; a invenção reivindicada não pode ser considerada nova e não pode ser considerada envolver uma atividade inventiva quando o documento é considerado isoladamente. "Y" documento de particular relevância; a invenção reivindicada não pode ser considerada envolver atividade inventiva quando o documento é combinado com um outro documento ou mais de um, tal combinação sendo óbvia para um técnico no assunto. "&" documento membro da mesma família de patentes.		
Data da conclusão da pesquisa internacional		Data do envio do relatório de pesquisa internacional:
6 October 2017		17/10/2017
Nome e endereço da ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Funcionário autorizado Vilella, Josep
Nº de fax:		Nº de telefone:

Formulário PCT/ISA/210 (segunda página) (Julho 2008)

RELATÓRIO DE PESQUISA INTERNACIONAL
Informação relativa a membros da família da patentes

Depósito internacional Nº

PCT/IB2017/053707

WO	2015198284	A1	30-12-2015	NONE		

US	2011188342	A1	04-08-2011	CN	101978424 A	16-02-2011
				EP	2255359 A1	01-12-2010
				JP	2011516830 A	26-05-2011
				KR	20100116223 A	29-10-2010
				US	2011188342 A1	04-08-2011
				WO	2009115299 A1	24-09-2009

Formulário PCT/ISA/210 (anexo da família de patentes) (Julho 2008)

Anexo 2 – Artigo publicado na conferência “*India's Digital Transformation*”

An Autonomous Wheelchair with Indoor Positioning System and Smart 3D Headphone for the Visually Impaired

João S. Pereira
Polytechnic Institute of Leiria, School of Technology and Management
Instituto de Telecomunicações, Portugal
Center for research in Informatics and Communications, e
Polytechnic Institute of Leiria,
Leiria, Portugal
joao.pereira@ipleiria.pt

Gurukiran Manjunath
Polytechnic Institute of Leiria, School of Technology and Management, Portugal
Bangalore, India
gkiran37@gmail.com

Nuno Almeida
Polytechnic Institute of Leiria, School of Technology and Management
Leiria, Portugal
nuno.ml.almeida@gmail.com

Sílvio P. Mendes
Polytechnic Institute of Leiria, School of Technology and Management
Center for research in Informatics and Communications,
Polytechnic Institute of Leiria,
Leiria, Portugal
smendes@ipleiria.pt

Abstract—Much research on electric-powered wheelchairs (EPWs) has been carried out in recent years to mitigate the problems of disabled people. We have come a long way since the basic manually-operated wheelchairs, and the progress made has considerably improved freedom of mobility of the disabled. The Internet of Things (IoT) is an upcoming area that could improve the autonomous operation of EPWs by providing independent maneuvering around obstacles. This study proposes an autonomous EPW having smart driving features for the visually impaired. The initial prototype of the autonomous EPW had voice, eye-movements, and GPS control. This “Instituto Politécnico de Leiria/Instituto de Telecomunicações” (IPL/IT)-based wheelchair has remote-control features that could benefit those with low-level visual disability. The tele-assistance features could assist the care-givers of the wheelchair user. Indoor locating systems using 3D printed directional Wi-Fi antennas allow for precise positioning of the wheelchair. The IPL/IT EPW system has been designed with low-cost IoT technology and, for this reason; it will be affordable to all. In addition, the EPW system has been enhanced with a novel smart 3D headphone that can generate acoustic signals in proportion to a 3D object, which would assist the visually impaired in identifying the shape of obstacles, and thus, avoid them.

Keywords—indoor positioning system, autonomous wheelchair, eye-controlled wheelchair, Internet of Things (IoT), telemedicine, tele-assistance, visually-impaired, seeing with 3D sounds.

I. INTRODUCTION

One of the most frequently utilised assistive devices for improving the personal movement of people with disabilities is the wheelchair. The availability of wheelchairs that are

suitable, well-made, and equipped not only improves mobility, but also opens the way to education, employment, and better prospects for disabled people [1]. Wheelchairs are thus an important part of daily lives as they can positively influence the quality of life for people with diminished mobility [2]. In most cases, manual or electric powered wheelchairs (EPW) are adequate. However, in instances where there is a lack of independent mobility, for example, in people with impaired vision, tremors, visual field reduction, spasticity, or intellectual impairment, wheelchair usage requires some assistance. Therefore, it can be seen that there is a requirement for smart wheelchairs which can independently plot a route to a desired destination without the intervention of the user [3]. A smart or autonomous wheelchair can be defined as “a uniquely modified powered wheelchair which is equipped with a control system and variant sensors” [4]. It can otherwise be called a robot base with an affixed seat that can move [5].

The objective of smart wheelchairs is to assist users who have some form of impairment that prevents them from navigating the wheelchair on their own, or who require the wheelchair to navigate independently to a chosen location, such as individuals with serious mobility impairments or who are visually impaired. Their main purpose is to reduce or eliminate the user's task in navigating the wheelchair. Moreover, their design needs to be customized to the user's condition and impairment.

There are two variants of smart wheelchairs: a regular power wheelchair with an added on-board computer and a collection of sensors, and a moveable robot base with an affixed seat.

The second version was the initial kind of smart wheelchair [5]. Presently, most commercially available smart wheelchairs are power wheelchairs which have been considerably modified, such as MAid (Mobility Aid for Elderly and Disabled people), NavChair, OMNI (Office wheelchair with high Maneuverability and Navigational Intelligence), and SENARIO [6–9]. Other smart wheelchairs such as Hephaestus, TinMan, Siamo, Smart Wheelchair Component System (SWCS), and Smart Power Assistance Module (SPAM) are designed as auxiliary units that can be joined to and removed from the base wheelchair infrastructure [10–14]. Generally, attention is required in two important areas for the design of an autonomous wheelchair. These are customisability and taking safety requirements into account [15]. Customisability indicates that the design of a smart wheelchair considers the needs of potential users, in general, and also incorporates the flexibility to be modified to suit the needs of a specific individual, in particular. Furthermore, designers should ensure that the safety of the wheelchair systems due to the disabilities of the potential users is maintained [15].

In general, an autonomous wheelchair's primary functions and suitable operational units are: user interface, navigation, and obstacle recognition and evasion [15]. Methods used for user interface in autonomous wheelchairs include various types of joysticks (e.g., standard, force-feedback, etc.), chin or head control, sip-and-puff devices, user facial expressions, and voice recognition [7, 8, 16–20]. Sensors are used to identify and evade obstacles. Different sensor types can be used for this purpose during the wheelchair's development. Multiple sensors may be used in a single wheelchair such as cameras for image processing, various kinds of range finders is maintained is maintained such as ultrasonic (sonar), infra-red (IR), and laser (LRF), bump sensors, and optical fibre gyroscopes [21–23]. Smart wheelchairs can vary in the manner of navigation assistance they offer, e.g. from the ability to follow targets to being limited to simple collision avoidance [8, 24]. Also, they may utilize an internal map to independently navigate to a specified destination. Tracks may also be utilized to facilitate autonomous navigation [25].

Many kinds of locating systems have been developed to estimate an autonomous mobile robot's absolute position [26] in an indoor set-up. In one study [27], a laser range finder was utilized to estimate a mobile robot's position which identified artificial landmarks positioned in the environment. Here, utilization of a costly sensor was a limitation. In another study [28], data gathered from the ultrasonic sensors of the robot was matched with an environmental global map for calculating its position. Despite having low-cost sensors, this technique required large-scale computations, including changes in the global map because of the variable nature of the environment. Additionally, studies have been conducted radio frequency (RF) operated systems for determining the location of mobile robots [29, 30]. In one such study, moving objects estimate their location by calculating the flight time of the signal received, called the Locate system [29]. In another

study, a Q-Track [30] system enabled the moving objects to send signals which were received by fixed receivers. This information is sent to a central unit which calculates the object locations. The effectiveness such as indoor systems is limited due to multiple reflections of the RF signal. The following sections describe the concept of Internet of Things used in this study and the design, implementation, and testing details of the proposed system. A novel headphone that converts reflected sound waves from obstacles into 3D acoustic signals has also been presented.

In October 2003, Blue Edge Bulgaria announced the development of a software that turns compatible phones into visual aids for the blind by changing images snapped by the camera into sounds that the user's brain can reconstruct into pictures [31]. Blue Edge's software is similar to the vVOICE system. The vVOICE system translates moving images into sounds in real-time. However, the Blue Edge's software also transforms the images. Once per second, the computer scans a 64x64 pixel frame. Each pixel in a column produces a wave whose frequency indicates its position; the highest frequencies are at the top. The sound waves are produced based on the 16-tone gray scale amplitudes of each pixel. Once all data has been extracted, the referred system grabs and digitizes a new video frame that orients the listener. To further boost the spatial orientation of the listener, stereo headphones are used to shift the volume balance from left-to-right in step with the movement of the pixel scanner. This gives the person a sense of the location of the objects. A low-resolution camera was chosen for capturing the images because the human ear has a lower capacity than the eye for handling data [32].

In 2014, a research team from a Brazilian University proposed a system that allows visually-impaired people to determine the trajectory of an object. Their work proposes a signification model that uses a transference function that receives the two-dimensional (2D) coordinate position of a tracked object as input and generates a synthesized sound as output. Their goal was to use sound to represent the trajectory described by an object to help visually-impaired persons who were not born to recognize simple geometric forms [33].

II. INTERNET OF THINGS

The IoT is a network of intelligent mechanisms interconnected to each other having unique digital identities [34, 35]. Items used daily are provided with an intelligent identity capable of detecting, comprehending, and responding to signals as a result of the union of new technologies and the Internet [36]. The IoT architecture comprises of three layers [34]:

- Perception layer: This layer has sensors for sensing and gathering data about the surroundings and locating smart devices in its surroundings.
- Network layer: This layer connects with neighboring network and smart devices and servers, processes sensor data, and transmits it.
- Application layer: This provides application-based utilities to the user.

The IoT system architecture proposed in this study is shown in Fig. 1.

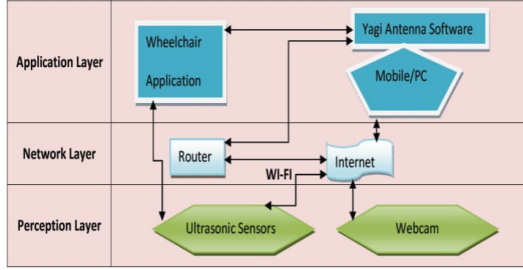


Figure 1. IoT architecture of proposed system.

III. BROWSER SECURITY

Hyper Text Transfer Protocol Secure (HTTPS) is a means to secure HTTP, the transfer protocol through which data is transferred between a browser and the connected website (the 'S' stands for 'Secure'). It indicates that all communications between the ends are encrypted. HTTPS is being used nowadays to protect confidential transactions taking place online, i.e. banking and online shopping order forms. Internet Explorer, Firefox and Chrome, additionally, in the address bar, show a padlock icon to indicate that a secure HTTPS connection is in place. Such pages generally use TLS (Transport Layer Security) or SSL (Secure Sockets Layer) encryption. Both protocols use an asymmetric Public Key Infrastructure (PKI) system which uses public and private keys for encrypting communications. Encryption and decryption is only possible with either of the keys. All system communications in this project are encrypted and secure.

IV. DESIGN AND IMPLEMENTATION OF PROPOSED SYSTEM

A. Wheelchair functionalities

The hardware components of the proposed IPL/IT autonomous EPW with easy to use parts is shown in Fig. 2 and Fig. 3. The primary system control is achieved by an economical Raspberry Pi minicomputer. It interacts well with other components and is flexible and performs numerical calculations well. The movement of the EPW is carried out by two DC motors with speed controllers. A simple joystick module steers the EPW. Speed, voice and eye motion instructions, barrier detection, and autopilot setting are controlled as specified in this section. The custom software, developed by a different IPL/IT group, manages all the peripheral sensing devices. The EPW uses control by ocular movements. The images, which are collected by a webcam, also assess user intent by reading eye motion. Two instruction sets are created for eye-movements: direction and halting of all movements. In addition, tracking ocular movements, validating voice commands, and autopilots functioning by Internet are carried out. Figure 4 presents our graphical user interface. A personalized voice recognition application (off-line) for voice

controlling is integrated with the motors and is managed by a computer. Instructions for directional motion and halting are managed by a voice command algorithm. A wheelchair that is controlled by voice helps a person with disabilities to be independently mobile. This can be achieved by interfacing a voice recognition application with the motors. As the wheelchair prototype uses a Raspberry Pi minicomputer, the DC motors can be controlled using such a voice recognition module to move in accordance to the voice commands. The voice command algorithm "Correlation of Spectrograms", applied to our IPL/IT wheelchair, was developed by João S. Pereira, in 2001.



Figure 2. Block diagram of proposed system.



Fig. 3. The low cost autonomous IPL/IT wheelchair system, with voice command, eyes movement detection, GPS, colored line follower, and sound reproducer of 3D objects scanned

The IPL/IT wheelchair system contains manual controls for Front, Left, Right, and Reverse movements. Additionally, the wheelchair has smart commands, such as reading eye movements, with a webcam fixed on a helmet; validation by voice commands using a microphone; autopilot using a Global Positioning System (GPS) or a colored line; and autopilot via Internet (Web). Moreover, the IPL/IT wheelchair system is low priced (lower than 400€) and is a modular system that can be modified to suit with any wheelchair. Furthermore, its controls are configurable to any user.



Fig. 4. Graphic interface of the eye-controlled system

The hardware design for the novel 3D surround headset system is illustrated in Fig. 5. The system, added to the IPL/IT wheelchair, comprises of (1) three pairs of speakers, (2) an image-based 3D scanner, and (3) an ultrasonic 3D scanner. The ©Microsoft Kinect 3D scanner is used when the object is visible to the user; else the ultrasonic 3D scanner is used. This new model has been incorporated in our IPL/IT autonomous wheelchair.



Fig. 5. 3D surround headphones schematics.

The Raspberry Pi minicomputer serves as the heart of the system and offers an interface between the units, namely the GPS, the collision avoidance sensors and other sensors. The

wheelchair unit receives latitude and longitude information through the GPS system. We display this information on Google Maps by using Google application programming interfaces (API), thus improving the user experience. Figure 6 illustrates our graphic user interface. A mobile app offers both an enriched experience to the user and handy information regarding the location of the wheelchair. The app communicates on a real-time basis with the server. We refresh the data (latitude and longitude) at intervals of 1 second. The positioning system also has the capability to regulate the movements of the wheelchair and its direction/orientation. This is moreover a process to improve the safety of the user. The data is registered and documented in a defined format.



Fig. 6. Locating Wheelchair on Google Map

Therefore, our location wheelchair system includes:

1. A web server running on Raspberry Pi,
2. A mobile phone which utilizes a browser to view the images from the webcams of the wheelchair, and
3. A server-provided web page which
 - i. permits dynamic commands for real-time control of the two motors of the wheelchair,
 - ii. offers supplementary information regarding the different sensors utilized to avert issues, such as collisions (ultrasonic sensors and a Kinect scanner), and
 - iii. Gives the GPS location of the wheelchair.

The GPS works well in an outdoor scenario, with accuracy around 5 meters. This precision is not good enough to drive a wheelchair autonomously. Additionally, it is well-known that the GPS was not designed to work inside a building. For this reason, a new indoor positioning system for our wheelchair is presented. Our new system was built with some Wi-Fi ESP8266 modules and a Raspberry Pi to drive directional motorized antennas which gather longitudinal and latitudinal coordinates using a triangulation approach described later. These IoT devices are low cost and provide an affordable solution to any IPL/IT wheelchair user.

Low cost Yagi antennas have been used in the present study since the use of an array of static antennas is expensive. The coordinates of at least two Wi-Fi Yagi antennas are then gathered utilizing a Raspberry Pi. Flores, Marcillo, and Pereira, in 2017, utilized a comparable system to assess the position of

a drone. The same positioning system is reused with the IPL/IT autonomous wheelchair. The wheelchair can be manipulated by sending commands to the main server from a mobile app. The server then passes these commands to the hardware actually connected to the Internet via the Wi-Fi network which then generates suitable signals to guide the wheelchair.

Figure 7 depicts our portable Wi-Fi Yagi antenna system printed in a 3D printer.

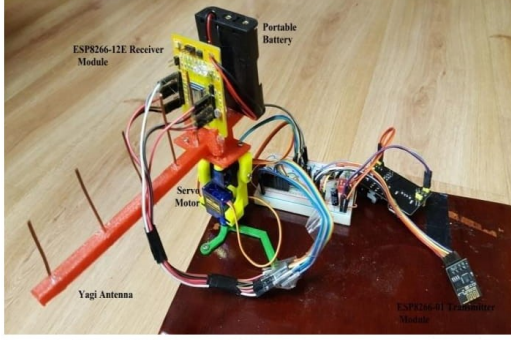


Fig. 7. Portable Wi-Fi Yagi antenna for an accurate indoor positioning system (prototype)

B. Triangulation process

The positional coordinates are arrived at using the Angle of Arrival (AoA) method (Fig. 8). This project suggests a novel localization method using Wi-Fi ESP8266 modules and Yagi motorized antennas. Initially, the angles θ_1 and θ_2 are measured for all nodes (here, P1 and P2) using the AoA method. Then, the line representing the orientation of highest propagation is calculated at all node references, their coordinates (X_i, Y_i) , $i=1, 2$ and angles $(\theta_1$ and $\theta_2)$ being known. The intersecting point of the lines $S(X, Y)$ is arrived at as follows [37]:

For line P_1S , slope $= \tan\theta_1 = \frac{Y-Y_1}{X-X_1}$,

we obtain:

$$Y = X \tan\theta_1 + Y_1 - X_1 \tan\theta_1 \quad (1)$$

Again, for line P_2S , slope $= \tan\theta_2 = \frac{Y-Y_2}{X-X_2}$.

Rearranging, we obtain:

$$Y = X \tan\theta_2 + Y_2 - X_2 \tan\theta_2 \quad (2)$$

Solving for X and Y , we get:

$$X = \frac{X_1 \tan\theta_1 - X_2 \tan\theta_2 + Y_2 - Y_1}{\tan\theta_1 - \tan\theta_2} \quad (3)$$

Similarly, we obtain:

$$Y = \frac{(X_1 - X_2) \tan\theta_1 \tan\theta_2 - Y_1 \tan\theta_2 + Y_2 \tan\theta_1}{\tan\theta_1 - \tan\theta_2} \quad (4)$$

Thus, the coordinates of S are:

$$S = \left(\frac{X_1 \tan\theta_1 - X_2 \tan\theta_2 + Y_2 - Y_1}{\tan\theta_1 - \tan\theta_2}, \frac{(X_1 - X_2) \tan\theta_1 \tan\theta_2 - Y_1 \tan\theta_2 + Y_2 \tan\theta_1}{\tan\theta_1 - \tan\theta_2} \right)$$

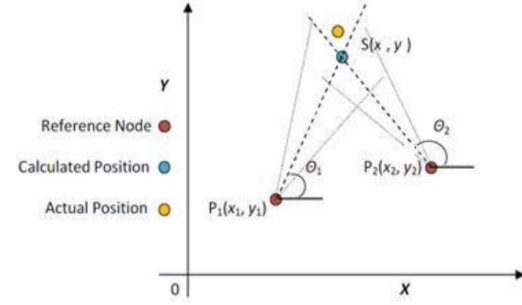


Fig. 8. Triangulation method using AoA.

The ESP8266 directional antennas are positioned on two servo motors which have a range of 180 degrees. In our scenario, the wheelchair is located inside a geographical equilateral triangle of 5 meters. The IoT technologies with 3 ESP8266 modules (Fig. 7) are placed at the vertices of the triangle. It is possible, in this manner, to discover the location of the wheelchair with a minimum error of few centimeters, much lower than the GPS error.

As mentioned earlier, the objective of the current initiative is to improve the present facilities of the IPL/IT wheelchair prototype with the intent of enhancing its user-friendliness (the Raspberry Pi is simple to operate), robustness, and accountability to enable both users and wheelchair providers to more successfully locate the wheelchair by means of a mobile application. This initiative uses IoT technologies to enhance the earlier IPL/IT wheelchair system. The primary elements of the extended system are: a centralized server with a Raspberry Pi minicomputer for control and a set of servo motors for each ESP8266 antenna. A network of ESP8266 modules was used to obtain the wheelchair location. All the modules are distributed in a real indoors scenario and they use Wi-Fi to communicate with the wheelchair server.

C. Multipath interference and standing waves

In wireless communications, multipath is a phenomenon of propagation in which radio signals arrive at the receiving antenna by more than one path. It results in multipath interference (MPI), which may be constructive and destructive interference, and causes the phase of the signal to get shifted. In this study, two Radio Frequency (RF) emitters (ESP8266 having Yagi antennas) are placed on the tables as illustrated in Fig. 9. Standing waves can be created when obstacles are placed in the path of a direct wave. The floor, for instance, can generate a standing wave by reflection of the direct wave. Similarly, a wheelchair or the walls, as obstacles, can generate many standing waves by causing a reflection of a direct wave. To overcome these issues, a positioning system which uses a

new method for cancelling standing waves [38] has been described in this study. This enhances the positional accuracy of obstacles and minimizes the effects of standing waves. In Fig. 9, the system attempts to accurately locate the wheelchair. The wheelchair has a Wi-Fi module that is used to identify the real location, by mean of a triangulation process. Each Wi-Fi directional antenna makes a scan of over 180 degrees until it discovers the direction of the Wi-Fi emitter of the wheelchair.

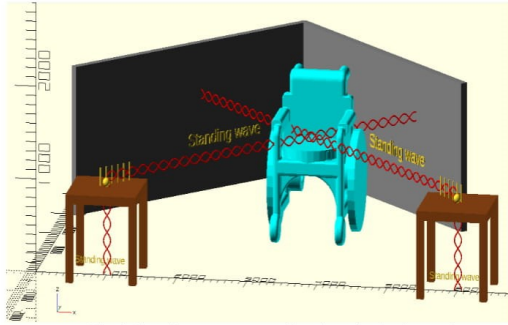


Fig. 9. Standing waves created by obstacles (scale: mm)

Figure 10 depicts a graph displaying the construction of a specific standing wave. The standing wave effect has a major impact on the accuracy of location estimation in a radio wave mechanism, due to the constant change in amplitude of the wave that affects signal strength readings, and thus, location accuracy. The standing wave is a wave with a full wavelength λ , where λ is also the wavelength of the RF carrier. To mitigate the effect of the standing waves, we use hardware-based methods to generate two different waves, in the pending patent [30]. The method generates a second wave with a variable delay that reduces the slow fluctuation of the standing wave.

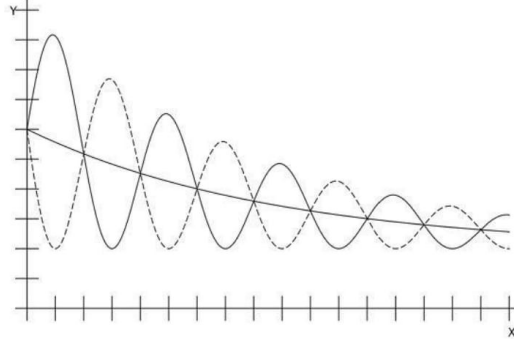


Fig. 10. Construction of a specific standing wave (X-axis is the distance and the Y-axis is the power of the signal received)

Figure 11 shows a method to reduce the standing waves effect and the RF shadow zone that can be useful on an estimation of indoor positioning system. The standing wave phenomenon has a detrimental influence on the estimation accuracy when radio waves are used since wave amplitude

varies constantly, thus affecting signal strength, and finally, the positional accuracy. To overcome the standing-wave effect, a hardware-oriented technique has been carried out to move the RF emitter in order to mitigate the standing wave effect, in our 3 pending patents [38–40]. In resume, it is a simple method that consists to move the initial location of an RF emitter around a circle of half wavelength ($\lambda/2$).

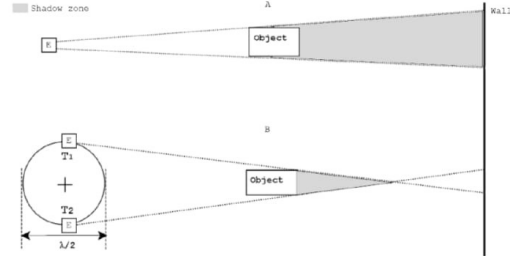


Fig. 11. Mitigation of a standing wave and reduction of RF shadow zone by moving the initial position of an emitter (E).

A 2D positioning system was set up for optimal transmission and receipt of data to study system accuracy and location error. Our device uses an automatic rotation of the Yagi antennas that performs a full scan of our indoor scenario, where the wheelchair is located.

As it can be seen in Fig. 9, the two ESP8266 RF emitters with Wi-Fi Yagi antennas was located at 5 meters. It is seen from Table 1 that the error, in the 3 different scenarios, increases with growth in antenna distances from the wheelchair. The average error of the three experiments was 30 cm. With the application of the patent pending technique [30] there was a 46% improvement of error location. By other words, the error is reduced 46%. Tables 1 and 2 present the outcome of the above experiment conducted in a 5 m x 5m environment in a closed room. The results clearly show that the pending patents technique [38] and [39] significantly improves the accuracy of the wheelchair positioning.

TABLE 1. INDOOR LOCATION ERROR RESULTS.

Initial position (X, Y) of the wheelchair [m]	Calculated starting position [m]	Initial error [m]	Final position [m]	Calculated final position [m]	Final error [m]	Average of the new error [m]	% of error location improvement after applying the pending patents [19] and [20]
(2.3, 2.7)	(2.59, 2.207)	0.57	(5.6, 5.85)	(5.15, 5.42)	0.49	0.30	46
(2.6, 2.9)	(2.85, 3.43)	0.58	(2.95, 2.71)	(3.11, 2.39)	0.35		
(2.2, 3)	(1.89, 1.13)	0.33	(1.99, 2.7)	(1.99, 2.76)	0.05		

TABLE 2. OUTDOOR LOCATION ERROR RESULTS.

Initial position (X, Y) of the wheelchair [m]	Calculated starting position [m]	Initial error [m]	Final position [m]	Calculated final position [m]	Final error [m]	Average of the new error [m]	% of error location improvement after applying the pending patents [19] and [20]

(1.72, 1.11)	(1.673, 1.105)	0.04	-	-	-	0.14	33%
(1.23, 0.87)	(1.155, 0.875)	0.095	(2.80, 2.13)	(2.863, 2.252)	0.069		
(1.39, 1.15)	(1.227, 1.115)	0.19	(2.11, 2.58)	(2.293, 1.599)	0.119		
(2.35, 3.53)	(2.145, 3.201)	0.387	(5.10, 4.80)	(5.03, 5.04)	0.25		

D. Sound conversion of 3D objects for the IPL/IT EPW

A new surround 3D headphones device is presented in this paper. It uses 3D sounds with a 3D headphone to reproduce a 2D image. By hearing 3D sounds, the user will know not only the shape of an object, but also the depth of it. Through the use of a 3D scanner, it is possible to create 3D objects of the environment facing the IPL/IT wheelchair. With our new method, the surface of the object is virtually covered by multiple virtual sound sources. For each virtual sound source, located on the surface of the object, a distance is calculated between the user and this source. These distances are calculated to simulate the locations of the various sound sources from the three-dimensional space that reach three pairs of speakers assembled in a new 3D headphone. The sound signals propagate ideally in a homogeneous medium transmission without obstacles, distortions and reflections. The 3D object is decomposed into multiple parallel front layers that are accessed/used sequentially, with a periodic scan performed from the nearest to the farthest layer. Different audible frequencies are used to determine each of the frontal layers. The curves of each layers (cutting the object) are represented by a limited number of points that are used to simulate the origin of the sound sources in a three-dimensional space.

The IPL/IT EPW uses a ©Microsoft Kinect scanner to capture 3D objects and convert them to 3D sounds. This converter is designed based on our pending patent [41]. All objects can be scanned and modeled by a decomposition of layers in a 3D Cartesian system (Fig. 12). The next step is to use images of the shape of the object in each layers (several parallel plans along the z-axis). Each layer is identified by an assigned specific set of audio frequencies. Every parallel layer is mapped on a 2D plan which contains the contour of the object. All such curves are identified by a collection of points that simulate an acoustic signal source in the 3D space. The virtual point locations of virtual sound sources are represented by 2D polar coordinates (radius = R and angle = B) centered on a horizontal line passing through the center of the three-dimensional object space and parallel to the X-Y plane. The points of virtual location of each sound source have equal audible frequencies whenever the R rays are equal, despite the angles B being different in the range 0–360°. The virtual points with high R are represented by low audible frequency range and points with low R are represented by higher audible frequencies. The user can estimate the object shape based on the hearing of an audible frequency proportional to the radius R on each plane.

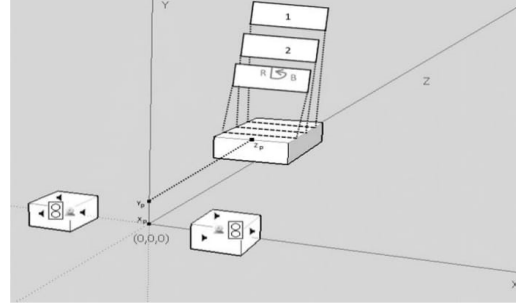


Fig. 12. 3D object decomposition, by layer, on a Cartesian system

To generate a particular sound for a particular speaker, the Euclidean separation between each point and each speaker p (x_p, y_p, z_p) is calculated as follows:

$$D = \sqrt{(x_p - x_{speaker})^2 + (y_p - y_{speaker})^2 + (z_p - z_{speaker})^2}$$

This distance D is used to generate the sound wave for the corresponding speaker using the formula for displacement $y(t)$ as:

$$y(t) = 256 * e^{-D} * \cos\left(\left(\frac{v}{D} * \frac{2\pi}{10^{-6} * D}\right) + (2\pi / (10^{-6} * D)) * \frac{t}{f}\right) \quad (7)$$

The experiments were carried out using the frequency $f = 44.1$ kHz. The velocity of sound in air was taken to be $v = 340$ m/s.

The software responsible for the generation of the sound starts by capturing the scenario or an object using a 3D Kinect scanner. After the 3D object is generated, it is split into various layers and the nearest layer is selected first. The outline of the object for that layer is represented in a 2D array, with the position of the array matching the outline of the object having a value (depth value) defined for that layer and the other positions having zero as value. The angle and radius of each sound source are calculated for each selected layer. After all, the angles and radius are calculated, the sound for each one of the 6 speakers (RB, RU, RD, LB, LU, LD) is generated (Fig. 13). When generation of all 6 sounds is finished, the left group of speakers plays the sounds generated for LB, LU and LD, while the right side plays the sounds for RB, RU and RD and then the software moves to the next layer. This process is repeated until the software reaches the farthest layer from the user. If, after all layers have been processed, our software detects any movement, this whole process will be repeated. Figure 13 can be used to better understand the previous method that it is described in the algorithm of Fig. 14.

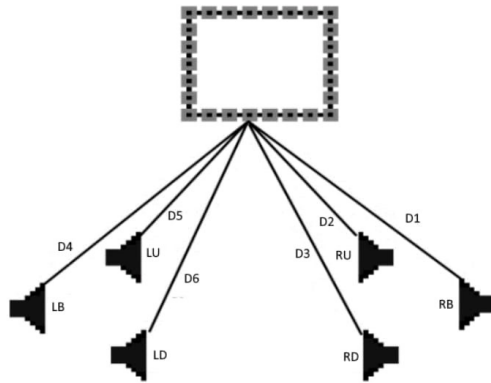


Fig. 13. Distance between speakers and points

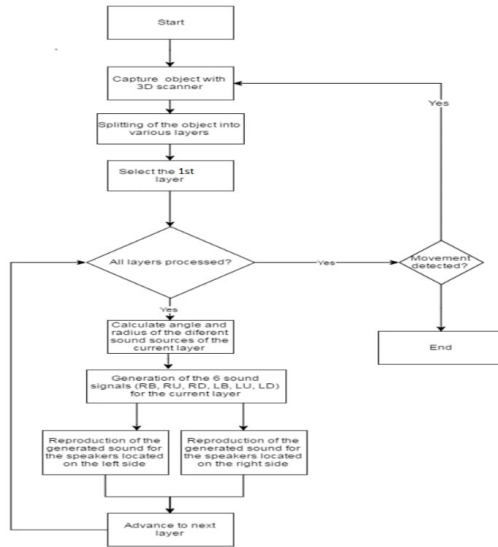


Fig. 14. Algorithm diagram

Figures 15–17 illustrate scans of a person making a hand gesture movement, a semi-closed door, and a closed door, respectively. The 3D objects are shown in the left, the scanned images converted into frequency sounds in the right (a different color for each frequency) and the generated sound amplitude below (delivered to 6 different pair of speakers). The colored points of the scanned images represent sounds of different frequencies, for all layers. Comparing these three figures, we observe that the more the variation in depth, the longer the duration of sound. After some training, using the 3D headphone of Figure 5, a blind person was able to easily recognize the three different objects. However, it may be possible to recognize other objects if intensive training is provided. In this

paper we proposed a solution to convert 3D objects into sound to give visually impaired people an idea of what an object looks like and how distant it is. Compared to other solutions that use 2D images, the use of a 3D object gives us a better understanding of that object. Instead of using the intensity of the pixels of an image to generate a sound to describe the image, our method (pending patent [33]) uses separate layers of the object to give visually impaired people a better understanding of the object's shape and depth.

To test this solution, a small group of 5 people was selected. The test consisted in playing the sounds of the closed and the semi-closed doors five times each with the person listening to the sound knowing to which object that sound corresponded to. After listening to all ten sounds, each person was given ten new sounds, five of a different closed door and five of a different semi-closed door, in a random order and had to guess which sound corresponded to which object. The test group managed to guess the right object 84% of the time.

With the tests performed, it was possible to identify not only the shape of the object, but also to identify its location in a 3D space. In the future, we intend to move the current computer application to a more portable solution. Along with this portable device, we also intend to take more tests with other real obstacles facing the autonomous IPL/IT wheelchair.

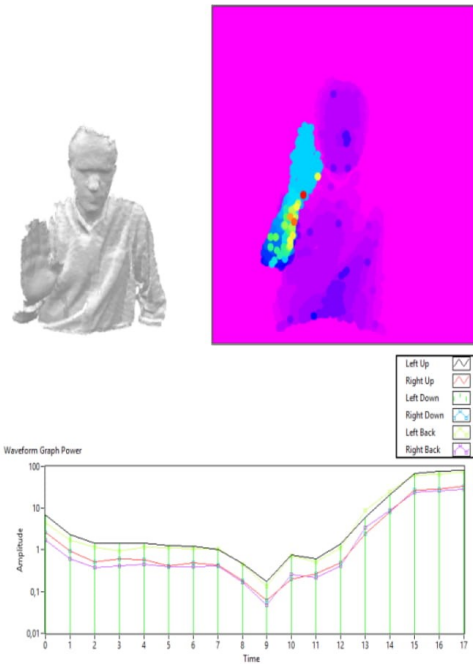


Fig. 15. Audible scan of a person

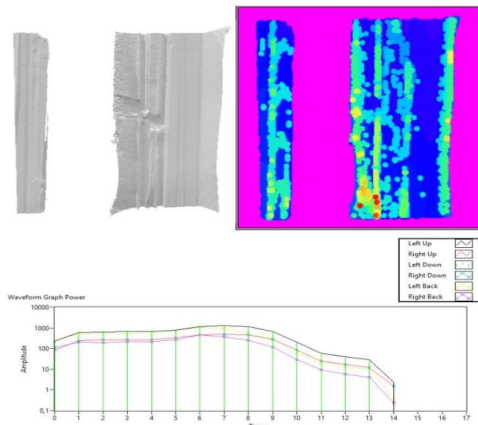


Fig. 16. Audible scan of a semi-closed door

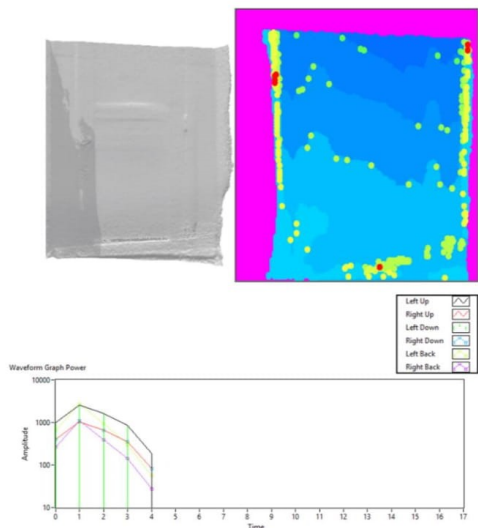


Fig. 17. Audible scan of a closed door

CONCLUSIONS

In this study, a smart autonomous wheelchair (IPL/IP EPW) having voice and eye-movement control has been described. The system uses a Wi-Fi network of ESP8266 modules to implement an accurate positioning system. This new system, based on some pending patents, has been tested successfully. The proposed system presents the location of a wheelchair with a positioning accuracy around 30 cm, when the experiment is performed within an indoor scenario. In addition, the IPL/IT EPW has been fitted with a novel 3D scanner that generates 3D sound appropriate to the obstacle sensed. This feature combined with the proposed positioning system will be beneficial to the visually challenged.

ACKNOWLEDGEMENT

The authors would like to thank the following students for their contributions, building the previous IPL/IT EPW prototype: Catarina Curioso, Duarte Ferreira, Flávia Simão, Inês Grenha, Mauro Pinheiro, Nuno Santos, Sílvia Fernandes, Tiago Neves, Guilherme Nunes, Fábio Matos, Nuno Domingues, Maurício Quaresma, Rui Paiva, Micael Florêncio, Ricardo Ferreira, João Santos, Christophe Silva, Virgílio Reis, Rui Vieira, João Silva, Elza Sousa, and Pedro Silva. Additionally, we would like to thank to the Professors Ricardo Martinho, Pedro Martinho, and Patrício Domingues. Finally, we thank the Polytechnic Institute of Leiria, School of Technology and Management of Leiria, and the "Instituto de Telecomunicações" of Portugal.

REFERENCES

- [1] World Health Organization [WHO]. (2017) *Guidelines on the provision of manual wheelchairs in less-resourced settings*. Available from: <http://www.who.int/disabilities/publications/technology/wheelchairguide-lines/en/> [Accessed 14th November 2017].
- [2] N. Thapar, G. Warner, M.L. Drainoni, S.R. Williams, H. Ditchfield, J. Wierbicky, and S. Nesathurai, "A pilot study of functional access to public buildings and facilities for persons with impairments," *Disabil. Rehabil.*, vol. 26, no. 5, pp. 280-289, 2004.
- [3] R. C. Simpson, "How many people would benefit from a smart wheelchair?," *J. Rehabil. Res. Dev.*, vol. 45, no. 1, pp. 53-72, Jan. 2008.
- [4] M.H. Alsibai and S.A. Manap, "A Study on Smart Wheelchair Systems," *Int. J. Eng. Technol. Sci.*, December 2015.
- [5] R. C. Simpson, "Smart wheelchairs: A literature review," *J. Rehabil. Res. Dev.*, vol. 42, no. 4, p. 423, 2005.
- [6] U. Borgolte, H. Hoyer, C. Bühler, H. Heck, and R. Hoelper, (1998) "Architectural concepts of a semi-autonomous wheelchair," *J. Intell. Robot. Sys.*, vol. 22, no. 3, 233-253, 1998.
- [7] N. Katevas, N. Sgouros, S. Tzafestas, G. Papakonstantinou, P. Beattie, J. Bishop, P. Tsanakas, and D. Koutsouris, "The autonomous mobile robot SENARIO: a sensor aided intelligent navigation system for powered wheelchairs," *IEEE Robot. Autom. Mag.*, vol. 4, no. 4, pp. 60-70, 1997.
- [8] R. C. Simpson, S. P. Levine, D. A. Bell, L. A. Jaros, Y. Koren, and J. Borenstein, "NavChair: An assistive wheelchair navigation system with automatic adaptation," *Assistive Technology and Artificial Intelligence Lecture Notes in Computer Science*, pp. 235-255.
- [9] E. Prassler, J. Scholz, and P. Fiorini, "A robotics wheelchair for crowded public environment," *IEEE Robot. Autom. Mag.*, vol. 8, no. 1, pp. 38-45, 2001.
- [10] M. Mazo, "An integral system for assisted mobility [automated wheelchair]," *IEEE Robot. Autom. Mag.*, vol. 8, no. 1, pp. 46-56, 2001.
- [11] D. P. Miller and M. G. Slack, "Design and testing of a low-cost robotic wheelchair prototype," *Auton. Robots*, vol. 2, no. 1, pp. 77-88, 1995.
- [12] R.C. Simpson, E.F. LoPresti, S. Hayashi, S. Guo, D. Ding, and R.A. Cooper, "Smart Power Assistance Module for manual wheelchairs. Technology and Disability: Research, Design, Practice and Policy," in *26th International Annual Conference on Assistive Technology for People with Disabilities (RESNA)*, June 2003.
- [13] R. Simpson, E. Lopresti, S. Hayashi, I. Nourbakhsh, and D. Miller, "The Smart Wheelchair Component System," *J. Rehabil. Res. Dev.*, vol. 41, no. 3b, p. 429, 2004.
- [14] R. Simpson, D. Poirat, and F. Baxter, "The Hephaestus Smart Wheelchair system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 10, no. 2, pp. 118-122, 2002.
- [15] A. Lankenau and T. Röfer, "Smart Wheelchairs - State of the Art in an Emerging Market," *KI*, vol. 14, no. 4, 37-39, 2000.
- [16] G. Bourhis and M. Sahnoun, "Assisted control mode for a smart wheelchair," in *Rehabil. Robot., 2007. ICORR 2007. IEEE 10th Int. Conf.*, 158-163, June 2007.
- [17] D. M. Brienza and J. Angelo, "A force feedback joystick and control algorithm for wheelchair obstacle avoidance," *Disabil. Rehabil.*, vol. 18, no. 3, pp. 123-129, 1996.

- [18] L. Fehr, W.E. Langbein, and S.B. Skaar, "Adequacy of power wheelchair control interfaces for persons with severe disabilities: A clinical survey," *J. Rehabil. Res. Dev.*, vol. 37, no. 3, pp. 353, 2000.
- [19] T. Felzer and R. Nordmann, "Alternative wheelchair control," in *Proc. 1st Int. IEEE-BAIS Symp. Res. Assistive Technol.*, vol. 7, 67–74, 2007.
- [20] H. Li, M. Kutbi, X. Li, C. Cai, P. Mordohai, and G. Hua, "An egocentric computer vision based co-robot wheelchair," in *IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2016, pp. 1829–1836, 2016.
- [21] P.C. Hsieh, "Autonomous robotic wheelchair with collision-avoidance navigation," M.S. thesis, Texas A&M Univ., 2008.
- [22] E. Prassler, J. Scholz, M. Strobel, and P. Fiorini, "An intelligent (semi-) autonomous passenger transportation system," in *Proc. 199 IEEE/IEEE/ISAI Int. Conf. Intell. Transport. Syst.*, pp. 374–379, Oct 1999.
- [23] R. Tang, "A Semi-autonomous Wheelchair Navigation System," M.S. Thesis, Univ. of Canterbury, 2012.
- [24] S. Hemachandra, T. Kollar, N. Roy, and S. Teller, "Following and interpreting narrated guided tours," in *2011 IEEE Int. Conf. Robot. Autom.*, 2011.
- [25] H. Wakaumi, K. Nakamura, and T. Matsumura, "Development of an automated wheelchair guided by a magnetic ferrite marker lane," *J. Rehabil. Res. Dev.*, vol. 29, no. 1, p. 27, 1992.
- [26] Borenstein, J., Everett, H.R., Feng, L., Lee, S.W., Byrne, R.H.: Where am I? Sensors and Methods for Mobile Robot Positioning. University of Michigan (1996)
- [27] Yazıcı, A., Parlaktuna, O.: Mobile robot localization using laser range finder and artificial landmark. In: Proceedings of IKECCO'2005, pp. 129–133 (2005)
- [28] Rencken, W.D.: Concurrent localization and map building for mobile robots using ultrasonic sensors. In: Proceedings of the 1993 IEEE/RSJ International Conference on Intelligent Robotics and Systems, pp. 2192–2197. Yokohama (1993)
- [29] Barnes, J., Cranenbroeck, J.V., Rizos, C., Pahwa, A., Politi, N.: Long term performance analysis of a new groundtransceiver positioning network (LocataNet) for structural deformation. In: FIG Working Week 2007. Hong Kong SAR (2007)
- [30] Q-Track: www.q-track.com. Accessed 18 Dec 2012 (2012)
- [31] S.A. Dallas Jr and A.J. Erickson, "U.S. Patent No. 4378569," Washington, DC: U.S. Patent and Trademark Office, 1983.
- [32] W.E. Langbein and L. Fehr, "Research device to preproduction prototype: a chronology," *J Rehabil Res Dev*, vol. 30, issue 4, pp. 436, 1993.
- [33] The Voice, "Augmented reality for the totally blind," September 2017, unpublished.
- [34] M.C. Domingo, "An overview of the Internet of Things for people with disabilities," *J. Netw. Comput. Appl.*, vol. 35, no. 2, 584–596, 2012.
- [35] J.P. Espada, O.S. Martínez, J.M.C. Lovelle, B.C.P. G-Bustelo, M.A. Álvarez, and A.G. García, "Modeling architecture for collaborative virtual objects based on services," *J. Netw. Comput. Appl.*, vol. 34, no. 5, pp. 1634–1647, 2011.
- [36] L. A. Amaral, F. P. Hessel, E. A. Bezerra, J. C. Corrêa, O. B. Longhi, and T. F. Dias, "eCloudRFID – A mobile software framework architecture for pervasive RFID-based applications," *J. Netw. Comput. Appl.*, vol. 34, no. 3, pp. 972–979, 2011.
- [37] Z. Shan and T.-S. P. Yum, "Precise Localization with Smart Antennas in Ad-Hoc Networks," *IEEE GLOBECOM 2007-2007 IEEE Global Telecommunications Conference*, 2007.
- [38] J. Pereira, J.B. Bagaric, and S. P. M. Mendes, "Standing Wave Cancellation – Wireless Transmitter, Receiver, System and Respective Method," *Patent 109137*, February, 2016, (Pending).
- [39] J. Pereira, M.G. Gasparovic, Pooja, and G. Manjunath, "Standing Wave Cancellation and Shadow Zone Reducing Wireless Transmitter, System and Respective Method and Uses," *Patent 109332*, April, 2016, (Pending).
- [40] J. Pereira, M.G. Gasparovic, and M. P. M. F. Ferreira, "Indoor Positioning System and Method," *Patent 109950*, March, 2017, (Pending).
- [41] J. Pereira and N. M. L. A. Almeida, "Método e Aparelho de Criação de um Cenário Tridimensional," *Patent 109485*, July, 2016 (Pending)

Anexo 3 - Artigo aceite na conferência “IEEE Sponsored 4TH International Conference for Convergence in Technology (I2CT) 2018”

An Autonomous Wheelchair Design with Smart Control for the Visually Impaired

João S. Pereira, IEEE Senior Member
Polytechnic Institute of Leiria, School of Technology and Management
Instituto de Telecomunicações, Portugal
Center for research in Informatics and Communications, e
Polytechnic Institute of Leiria,
Leiria, Portugal
joao.pereira@ipleiria.pt

Gurukiran Manjunath
Polytechnic Institute of Leiria, School of Technology and Management, Portugal
Bangalore, India
gkiran37@gmail.com

Nuno Almeida
Polytechnic Institute of Leiria, School of Technology and Management
Leiria, Portugal
nuno.ml.almeida@gmail.com

Silvio P. Mendes
Polytechnic Institute of Leiria, School of Technology and Management
Center for research in Informatics and Communications,
Polytechnic Institute of Leiria,
Leiria, Portugal
smendes@ipleiria.pt

Abstract: A considerable research and development on wheelchairs has been conducted to meet the needs of the disabled. Since the early manual wheelchairs to their more recent electric powered counterparts, advancements have focused on improving autonomy in mobility. Integration of Internet of things (IoT) is a promising area that has been studied to enhance the independent operation of the electrical wheelchairs by enabling autonomous navigation and obstacle avoidance. This paper describes the design of an autonomous wheelchair of the IPL/IT with smart driving features for persons with visual impairments. The objective is to improve the current prototype of the intelligent wheelchair. The first prototype of the wheelchair was built to control it by voice, ocular movements, and GPS (Global Positioning System). Furthermore, the IPL/IT wheelchair acquired a remote-control feature which could prove useful for people with low levels of visual impairment. This tele-assistance mode will be helpful to the family of the wheelchair user or, simply, to a health care assistant. Indoor and outdoor positioning systems, with 3D printed directional Wi-Fi antennas, have been deployed to enable a precise location of our wheelchair. The underlying framework for the wheelchair system is the IPL/IT low cost autonomous wheelchair prototype that is based on IoT technology for improved affordability. Additionally, the wheelchair has been improved with a new intelligent headphone that can reproduce sounds proportionally to a 3D image, which helps visually impaired people recognize the shape of objects and avoid obstacles.

Keywords—indoor positioning system, autonomous wheelchair, eye-controlled wheelchair, Internet of Things (IoT),

telemedicine, tele-assistance, visually-impaired, seeing with 3D sounds.

1. Introduction

The demand for intelligent products for the care of the elderly, ageing, disabled, and infirm is steadily growing. A device, like an electric powered wheelchair (EPW), is an important movement aid for the disabled, providing self-reliance, mobility, and a better quality of life [1]. In 2010, there were some 540,000 United States of America citizens who had benefitted from utilizing EPWs, and this figure has continued to grow since [2]. The EPW is essential for patients with limited or nil movements of lower or upper-limbs, particularly those affected with spinal cord injuries, muscular dystrophy, cerebral palsy, etc [4, 5]. To aid disabled people, many EPWs are currently available in the market [6, 7]. They are of diverse kinds, and help the user by tilting, reclining, among other features [8]. Several control features for EPWs have been proposed in the literature, e.g. voice[9], facial orientation[10], ocular gaze[6], electromyography inputs from the muscles of the arms and neck [11], electromyography inputs from the facial muscles [12], ocular-control electro-oculography (EOG) method [13, 14], etc.

The present work describes an improvement to an autonomous wheelchair designed by Instituto Politécnico de Leiria/Instituto de Telecomunicações (IPL/IT) having smart features for the visually impaired. The earlier prototype was designed for control by ocular movements, voice, and Global Positioning System (GPS), along with a remote control for people having slight vision impairment. The system uses low cost automation based on the Internet of Things (IoT) technology.

2. Internet of Things

The concept of IoT refers to a network of smart devices in which they are all connected to each other and possess digital identities [15, 16]. Objects of daily use are given a smart identity with the ability for signal detection, comprehension, and response owing to the synergy between the new technologies and the Internet [17]. Domingo (2012) states that the architecture of the IoT is split into three layers [15]:

- A physical perception layer having objects like sensors for detecting and collecting information about the surroundings and locating smart devices in its surroundings,
- A network layer, which connects to nearby smart devices, servers, and network devices, and transmits and processes sensor information, and
- An application layer which delivers application-based utilities to the user.

The proposed IoT system architecture of this study is illustrated in Fig.1.

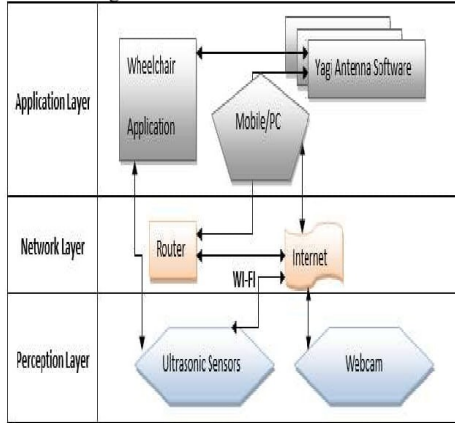


Figure 1. IoT architecture of proposed system.

3. Design and Implementation of Proposed System

3.1. Hardware Design

The proposed hardware design of the IPL/IT autonomous wheelchair with user-friendly components is illustrated in Fig. 2. The principal control of the proposed system is carried out by a low-cost minicomputer, a Raspberry Pi. It can interact with other devices and offers good flexibility and performance with numerical calculations. The motion of the wheelchair is performed by two DC motors having speed limiters. A joystick module controls wheelchair motion. There are components included for motor speed, voice and ocular motion commands, barrier sensing, and autopilot mode. The dedicated software, developed 4 years ago by another IPL/IT team, regulates all the peripheral sensors.

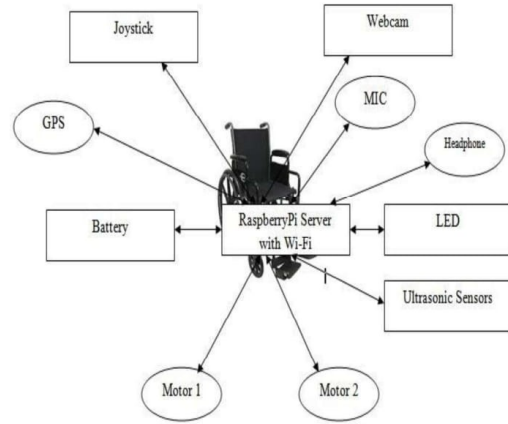


Figure 2. Block diagram of proposed system.

For the wheelchair's voice control feature, a voice recognition application is interfaced with the motors and controlled by a Raspberry Pi minicomputer. Commands for directional movements and stopping are regulated by a voice command algorithm. The wheelchair also employs eye movement control, whose images are gathered by a webcam in addition to assessing user intent by monitoring eye movements. Two kinds of commands are generated for eye control: directional movement and stopping of any movement. Additionally, reading eye movements, voice command validations, and autopilot functions by Internet are performed. Our new system proposal performs positioning using three Wi-Fi ESP8266 modules and a Raspberry Pi with motors to operate the directional antennas which obtain latitudinal and longitudinal coordinates by a triangulation process described in the next section.

The hardware design for the new 3D surround headphones system can be seen in Fig. 3. This system is composed by 3 pairs of speakers (1), an image-based 3D scanner (2) and an ultrasonic 3D scanner (3). The image-based 3D scanner is used when there is visual contact between the user and an object. The scanner chosen for this purpose was a ©Microsoft Kinect. When it is not possible to provide visual contact, the ultrasonic 3D scanner is used.



Figure 3. 3D surround headphones schematics.

3.2. Triangulation process

The positional coordinates are arrived at using the Angle of Arrival (AoA) method (Fig. 4). This project suggests a novel localization method using Wi-Fi ESP8266 modules and Yagi motorized antennas. Initially, the angles θ_1 and θ_2 are measured for all nodes (here, P1 and P2) using the AoA method. Then, the line representing the orientation of highest propagation is calculated at all node references, their coordinates (X_i, Y_i) , $i=1, 2$ and angles $(\theta_1$ and $\theta_2)$ being known. The intersecting point of the lines $S(X, Y)$ is arrived at as follows [34]:

For line P₁S, slope = $\tan\theta_1 = \frac{Y-Y_1}{X-X_1}$,

We obtain:

$$Y = X \tan\theta_1 + Y_1 - X_1 \tan\theta_1 \quad (1)$$

Again, for line P₂S, slope = $\tan\theta_2 = \frac{Y-Y_2}{X-X_2}$.

Rearranging, we obtain:

$$Y = X \tan\theta_2 + Y_2 - X_2 \tan\theta_2 \quad (2)$$

Solving for X and Y , we get:

$$X = \frac{X_1 \tan\theta_1 - X_2 \tan\theta_2 + Y_2 - Y_1}{\tan\theta_1 - \tan\theta_2} \quad (3)$$

Similarly, we obtain:

$$Y = \frac{(X_1 - X_2) \tan\theta_1 \tan\theta_2 - Y_1 \tan\theta_2 + Y_2 \tan\theta_1}{\tan\theta_1 - \tan\theta_2} \quad (4)$$

Thus, the coordinates of S are:

$$S = \left(\frac{X_1 \tan\theta_1 - X_2 \tan\theta_2 + Y_2 - Y_1}{\tan\theta_1 - \tan\theta_2}, \frac{(X_1 - X_2) \tan\theta_1 \tan\theta_2 - Y_1 \tan\theta_2 + Y_2 \tan\theta_1}{\tan\theta_1 - \tan\theta_2} \right) \quad (5)$$

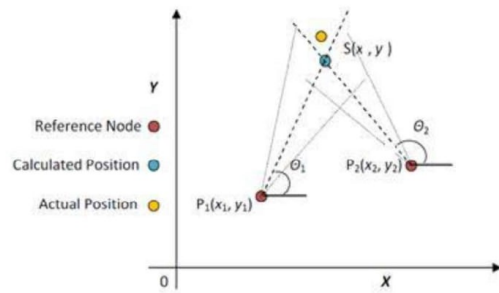


Fig. 4. Triangulation method using AoA.

The ESP8266 directional antennas are positioned on two servo motors which have a range of 180 degrees. In our scenario, the wheelchair is located inside a geographical

equilateral triangle of 5 meters. The IoT technologies with 3 ESP8266 modules (Fig. 7) are placed at the vertices of the triangle. It is possible, in this manner, to discover the location of the wheelchair with a minimum error of few centimeters, much lower than the GPS error.

As mentioned earlier, the objective of the current initiative is to improve the present facilities of the IPL/IT wheelchair prototype with the intent of enhancing its user-friendliness (the Raspberry Pi is simple to operate), robustness, and accountability to enable both users and wheelchair providers to more successfully locate the wheelchair by means of a mobile application. This initiative uses IoT technologies to enhance the earlier IPL/IT wheelchair system. The primary elements of the extended system are: a centralized server with a Raspberry Pi minicomputer for control and a set of servo motors for each ESP8266 antenna. A network of ESP8266 modules was used to obtain the wheelchair location. All the modules are distributed in a real indoors scenario and they use Wi-Fi to communicate with the wheelchair server.

3.3. Multipath interference and standing waves

In the presence of multiple obstacles, waves being transmitted travel through space and reflect off the various obstacles causing multipath interference (MPI) due to detecting several instances of the same wave. In our scenario, two radio frequency emitters (ESP8266 with Yagi antenna) are on the tables depicted in Fig. 5. Some standing wave can be generated in the presence of some obstacles in front of a direct wave. For example, the ground can reflect a direct wave and can create a standing wave. By the same way, an obstacle like a wheelchair can reflect a direct wave and generate more standing waves. To avoid these problems, a locating system that uses a novel standing-wave cancelation method [19] has been utilized. This method improves the positional accuracy locations of some objects and mitigates the standing wave effect. The object of Fig. 5 that we want to locate accurately is the wheelchair.

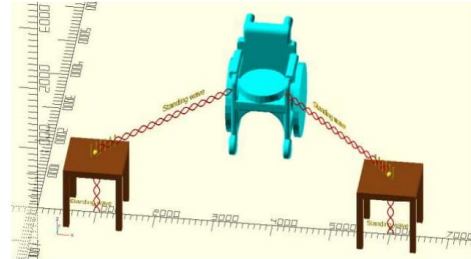


Figure 5. Standing waves generated due to the presence of some obstacles (millimeter scale).

Figure 6 depicts a graph displaying the effect of the standing wave on the indoor positioning system estimation. The

graph also displays the standing wave effect mitigation used within the system. The standing wave effect has a major impact on the accuracy of location estimation in a radio wave mechanism, due to the constant change in amplitude of the wave that affects signal strength readings, and thus, location accuracy. To mitigate the effect of the standing waves, we use hardware-based methods to generate two different standing waves, in the pending patent [19]. The first standing wave is a wave with a full wavelength of λ . The second reflected wave also has a wavelength of λ but starts with a shift of $\lambda/2$. The curve shown in Fig. 6 obtained from summing up the two waves proves to be better for standing wave cancellation when location estimation is considered.

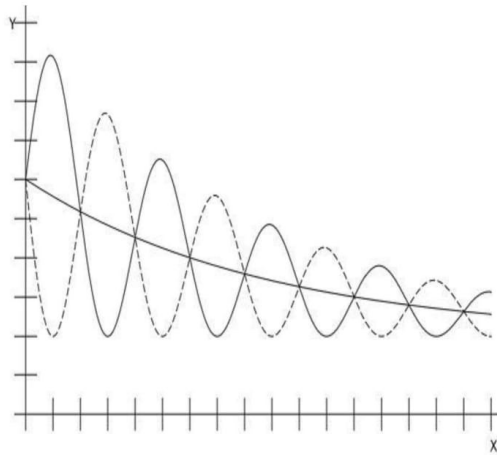


Figure 6. Construction of a standing wave (X -axis is the distance and the Y -axis is the power of the signal received).

A two-dimensional (2D) location system was considered to optimize the sending and receiving of data, and to have a better perception of the signal and accuracy of the system with almost no error value in location. This system also helps to reduce the location error of the devices that are both static (manual rotation of the Yagi antennas) and dynamic (automatic rotation/scan of the Yagi antennas) employing variable power strength. Table 1 shows the error results calculated in static and dynamic scenarios. It is understood that the error increases according to the distance of the nodes to the wheelchair. The average error in our experiments was 14.6 cm and the percentage of error location improvement was 33%, when our pending patented method [19] is applied.

horizontal line passing through the center of the 3-D

Table 1. Location error results.

State	Initial position (x,y) [m]	Calculated starting position (m)	Initial error (m)	Final position (m)	Calculated final position (m)	Final error (m)	%	Average of error	% of error location improvement
Dyna mic	(1.2, 0.8)	(1.1, 0.8)	0.09	(2.8, 2.1)	(2.8, 2.2)	0.06	27.1	0.14	33
Static	(1.3, 1.1)	(1.2, 1.1)	0.1	(2.1, 2.5)	(2.2, 1.5)	0.1	37.1		
Dyna mic	(2.3, 3.5)	(2.1, 3.2)	0.3	(5.1, 4.8)	(5.0, 5.0)	0.2	35		

It can be seen from Table 1 that the pending patent method [19] improves significantly the accuracy of the wheelchair location.

3.4. Sound conversion of a 3D object for the IPL/IT wheelchair

The IPL/IT wheelchair has a ©Microsoft Kinect scanner in front of it that catches 3D objects and converts them into 3D sounds. This device is based on our pending patent [20]. Any scanned object can be modeled on a three-dimensional Cartesian system (Fig. 7). The object is then decomposed in multiple parallel layers along the z -axis. Different audible frequencies are used to determine each of the layers.

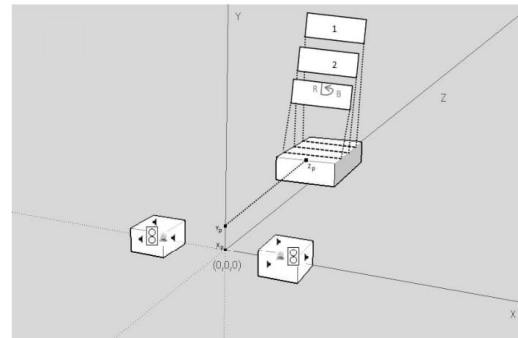


Figure 7. 3D object on the Cartesian system

Each one of these parallel layers is represented in a 2D plane, containing the outline/contour of the object. The curves of each plane are represented by a limited number of points that are used to simulate the origin of a sound source in a three-dimensional space. The points of virtual locations of virtual sound sources are represented by 2D polar coordinates (radius = R and angle = B) centered on a

object space and parallel to the z-axis. The points of virtual location of each sound source have equal audible frequencies whenever the R rays are equal, despite the angles B that can be different in the range $[0^\circ, 360^\circ]$. The virtual points with high radii R are represented by low audible frequency range and points with low rays are represented by higher audible frequencies. The user can estimate the object shape based on the hearing of an audible frequency proportional to the radius R on each plane. To create a specific sound for a specific speaker, the Euclidean distance between each speaker and each point $p(x_p, y_p, z_p)$ is calculated:

$$D = \sqrt{(x_p - x_{speaker})^2 + (y_p - y_{speaker})^2 + (z_p - z_{speaker})^2} \quad (8)$$

his distance D is used to generate the sound wave for the corresponding speaker using the formula:

$$y(t) = 256 * e^{-D} * \cos\left(\left(\frac{v}{D} * \frac{2\pi}{10^{-6} * D}\right) + (2\pi / (10^{-6} * D)) * \frac{t}{f}\right) \quad (9)$$

Our tests were made with the frequency $f = 44100$ Hz, the sound velocity/propagation in the air $v = 340$ m/s (t is the time in seconds).

Figure 8 shows the result of the scan of a person making a hand signal (blue dot area). Figure 9 presents the scan of a closed door. On both figures it is possible to see the scanned object (on the left), the points detected in the various layers (on the right) and the graphic of the amplitude of the generated sound for each speaker (below). Comparing the difference in both graphics it is possible to see that the higher the variation in depth, the longer the sound will play. Each colored point represents a different audible frequency

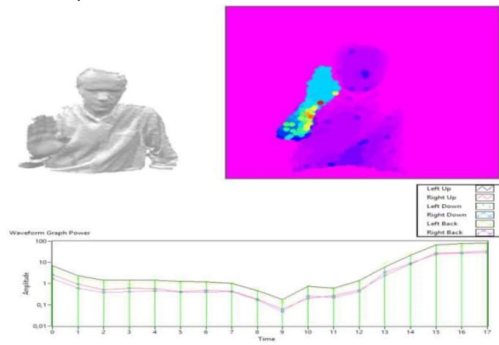


Figure8. Audible scan of a person

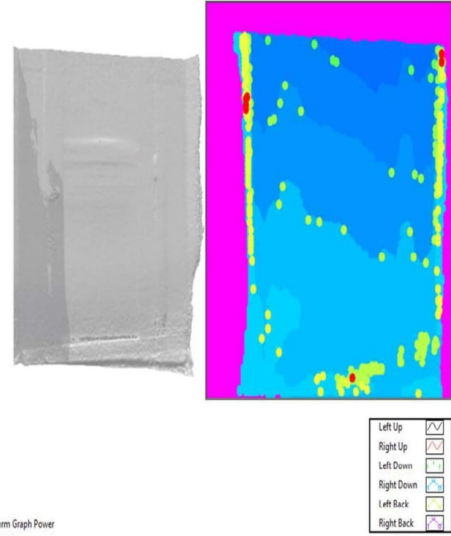


Figure9. Audible scan of a closed door

4. Conclusions and future work

In this paper, we have described an intelligent autonomous wheelchair controlled by voice and ocular movement which utilizes a small Wi-Fi network of ESP8266 modules to attain an accurate location. We have tested the system and have observed that the results are promising. It has been possible to locate a wheelchair that has a Wi-Fi emitter within an accuracy of 15 cm. Additionally, our IPL/IT wheelchair has been equipped with a novel 3D scanner that generates 3D sound based on each obstacle detected. This new ability combined with our new accurate positioning system will be useful to visually impaired.

In the future, we will look into implementing a more portable solution and to perform varied tests in real time, in different outdoor and indoor scenarios, with different wheelchairs simultaneously.

References

- [1] Subhadra Evans, Claudius Neophytou, Lorraine De Souza, and Andrew O. Frank. Young people's experiences using electric powered indoor-outdoor wheelchairs(EPIOCs): potential for enhancing users' development? *Disabil. Rehabil.* 2007; 29(16):1281–94.
- [2] Brault MW. Americans with disabilities: 2010. Current population reports [Internet]. Washington (DC): U.S. Census Bureau; 2012 (cited 2015 May 24; updated 2012 Jul). Available from: <http://www.census.gov/prod/2012pubs/p70-131.pdf>
- [3] Ding D, Cooper RA. Electric powered wheelchairs. *IEEE Control Syst.* 2005;25(2):22—3
- [4] Anderson KD. Targeting recovery: Priorities of the spinal cord-injured population. *J Neurotrauma.* 2004;21(10): 1371–83.
- [5] Murphy KP, Molnar GE, Lankasky K. Employment and social issues in adults with cerebral palsy. *Arch Phys Med Rehabil.* 2000;81(6):807–11.
- [6] Kumar Vijay, Assistive Devices for People with Motor Disabilities, Wiley Encyclopedia of Electrical and Electronics Engineering. Assistive Devices for People With Motor Disabilities, 1997.
- [7] Pires G., Autonomous Wheelchair for Disabled People, Proc. IEEE Int. Symposium on Industrial Electronics (ISIE97), Guimarães, 797-801, 1997.
- [8] R.F. Edlich, K.P. Nelson, M.L. Foley, R.M. Buschbacher, W.B. Long and E.K. Ma, Technological advances in powered wheelchairs, J. Long. Term. Eff. Med. Implants, vol. 14, pp.107-130, 2004.
- [9] Pires, G. & Nunes, U., "A wheelchair steered through voice commands and assisted by a reactive fuzzy-logic controller", *Journal of Intelligent and Robotic Systems*, Vol. 34, No. 3, pp. 301–314, 2002.
- [10] Bergasa, L.M., Mazo, M., Gardel, A., Barca, R. & Boquete, L., "Commands generation by face movements applied to the guidance of a wheelchair for handicapped people", *Proc. of the 15th International Conference on Pattern Recognition*, Vol. 4, pp. 4660–4663, 2000.
- [11] Matsumoto, Y., Ino, T. & Ogasawara, T., "Development of intelligent wheelchair system with face and gaze based interface", *Proc. of the 10th IEEE International Workshop on Robot and Human Interactive Communication*, pp. 262–267, 2001.
- [12] Kim, K.H., Kim, H.K., Kim, J. S., Son, W. & Lee, S.Y., "A bio signal-based human interface controlling a power-wheelchair for people with motor disabilities", *Etri Journal*, Vol. 28, No. 1, pp. 111–114, 2006.
- [13] Barca, R., Boquete, L., Mazo, M. & Lopez, E., "Wheelchair guidance strategies using EOG", *Journal of Intelligent and Robotic Systems*, Vol. 34, No. 3, pp. 279–299, 2002.
- [14] Rebsamen, B., Burdet, E., Guan, C., Zhang, H., Teo, C.L., Zeng, Q., Laugier, C. & Ang Jr., M. H., "Controlling a wheelchair indoors using thought". *IEEE Intelligent Systems*, vol. 22, No. 2, pp. 18–24, 2007.
- [15] Domingo, M. C. (2012) An overview of the Internet of Things for people with disabilities. *Journal of Network and Computer Applications*, 35(2), 584-596.
- [16] Espada, J. P., Martinez, O. S., Lovelle, J. M. C., G-Bustelo, B. C. P., Álvarez, M. Á. & García, A. G. (2011) Modeling architecture for collaborative virtual objects based on services. *Journal of Network and Computer Applications*, 34(5), 1634-1647.
- [17] Amaral, L. A., Hessel, F. P., Bezerra, E. A., Corrêa, J. C., Longhi, O. B. & Dias, T. F. (2011) eCloudRFID—Amobile software framework architecture for pervasive RFID-based applications. *Journal of Network and Computer Applications*, 34(3), 972-979.
- [18] Shan, Z. and Yum, T. S. P. (2007) Precise localization with smart antennas in ad-hoc networks. *Proceedings of the IEEE GLOBECOM'07*, Washington, DC, Nov. 2007, pp. 1053–1057.
- [19] J. Pereira, M.G. Gasparovic, Pooja, Gurukiran Manjunath, Standing Wave Cancellation and Shadow Zone Reducing Wireless Transmitter, System and Respective Method and Uses, 109332, April, 2016, (Pending Patent).
- [20] J. Pereira, N. M. L. A. Almeida, MÉTODO E APARELHO DE CRIAÇÃO DE UM CENÁRIO TRIDIMENSIONAL, 109485, July, 2016, (Pending Patent)